

© Copyright by

Sa A. Bui

May 2012

ESSAYS ON APPLIED MICROECONOMICS

A Dissertation

Presented to

The Faculty of the Department

of Economics

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

By

Sa A. Bui

May 2012

ESSAYS ON APPLIED MICROECONOMICS

Sa A. Bui

APPROVED:

Scott A. Imberman, Ph.D.
Committee Co-Chair

Aimee Chin, Ph.D.
Committee Co-Chair

Steven G. Craig, Ph.D.

Jonathan Meer, Ph.D.
Texas A&M University

John W. Roberts, Ph.D.
Dean, College of Liberal Arts and Social Sciences
Department of English

ESSAYS ON APPLIED MICROECONOMICS

An Abstract of a Dissertation

Presented to

The Faculty of the Department

of Economics

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

By

Sa A. Bui

May 2012

Abstract

This dissertation consists of three essays. In the first essay, I study the interactions of students with limited English proficiency (LEP). It is vital to study these LEP students' (LEPs) interactions because immigration into the United States continues to usher many foreign students with LEP into American schools. Additionally, U.S.-born children from immigrant homes enter schools knowing little English. In order to understand how best to instruct these limited English proficient (LEP) students, it is important to examine how concentration of LEP students (LEPs) affects educational outcomes of the LEPs themselves. On one hand, having a larger number of LEPs allows teachers to deliver a more focused instruction. On the other hand, LEPs may speak in their native languages more often and practice speaking English less when they are surrounded by many other LEPs. In this paper, I examine the effects of classmate English proficiency on the educational outcomes of LEP 5th graders using administrative data from an urban school district. Specifically, I study how exposure to LEPs affects the achievement, mainstreaming and grade retention of LEPs using the idiosyncratic variation in LEP shares across cohorts in a school. I find having more LEPs in a cohort leads to higher math achievement, faster mainstreaming and less grade retention amongst LEP students.

In the second essay, I partnered with Scott Imberman and Steven Craig. We identify the impact of gifted and talented services on student outcomes by exploiting a discontinuity in eligibility requirements and find no impact on standardized test scores of marginal students even though quality of peers and classes improve substantially. We then use randomized lotteries to examine the impact of attending a GT magnet program, relative to programs in other schools, and find that, despite exposure to higher quality teachers and peers, only science achievement improves. We find that the relative ranking of the students change, as do their grades, indicating that either invidious comparison peer effects or teaching targeting may be important.

In the third essay (with Adriana Kugler), we examine the impact of remittances on households' investments and consumption in Vietnam using the Living Standards Surveys. Given that households likely face budget constraints in Vietnam, one may expect for remittances to affect the decisions of households to invest and consume. In addition, since the unitary model of the household is particularly unlikely to represent households in developing countries, we also look at differential impacts when women receive a larger fraction of the remittances. We use an instrumental variables strategy to address the fact that households receiving different amounts of remittances and sending different amounts of remittances to women are likely to differ in terms of their observable and unobservable characteristics that correlate with investments and spending. We instrument the amount of remittances and the share of remittances going to women with the 1992 migration rate from the household's region of residence and the interaction between this variable and the share of women in the household. OLS results show that remittances are associated with better health of young, adult and older individuals, while the fraction of remittances received by women is associated with greater educational attainment and attendance, and less child labor while changing the composition of consumption expenditures from all categories towards health expenditures. However, when we use an IV strategy, we find that remittances increase education expenditure while reduce food expenditure. More importantly, the fraction of remittances received by women increases the household expenditure on health relative to other household expenditures. The results thus show not only the amount of remittances but also the identity of the receiver matters in terms of increasing human capital investments for children and their family members.

Acknowledgments

There are so many people whose support I have relied on throughout the years. I would not be where I am today without their presence in my life. First, I would like to thank my family for their love and understanding. Particularly, I want to thank my mom who has helped me with taking care of my daughter during my graduate school. I also want to thank my best friend Lan Anh Tran for her encouragement regardless of my frequent complaints. Lastly, I thank my husband, Richard Nguyen, for his patience, love, and unwavering support.

Another group of people I want to give thanks to are my teachers. They are the people who always believed in my potential even when I had little confidence in myself. First and foremost, I want to thank Mr. Jesse Alred, who was my high school U.S. Government and Economics teacher. It was Mr. Alred, who instilled in me an interest for public policy and a desire for change. Next, I want to thank Dr. Steven Craig, who was my Public Finance professor when I was an undergraduate student, and whose teaching inspires students to ask their own questions and search for their own answers. It was then that I wanted to be a professor, like Dr. Craig, to open up minds like he did mine. I want to thank Dr. Charles Becker, who presented me an opportunity to think about graduate school in economics. I want to thank my committee co-chairs, Dr. Scott Imberman and Dr. Aimee Chin. They introduced me to the possibility of being a researcher. I am grateful to have the opportunity to learn from their experience as researchers and teachers. I am inspired to be a scholar who has an eye for data like Dr. Imberman, who is critical like Dr. Craig, and who is detailed-oriented like Dr. Chin. I want to thank my other teachers and professors as well. To list all their names here would take many more pages. Additionally, I thank Dr. Jonathan Meer for agreeing to be on my dissertation committee. I appreciate his time and especially, his drive from College Station to Houston.

To those whose constant presence was an encouragement in itself, you are my friends and colleagues. Also, thank you to the staff at the Economics department and students at the University of Houston. I thank you for many fruitful and joyous conversations. I have benefited much from speaking to you.

to Lieu Pham & Kim Bui

Contents

1. How do Limited English Proficient Students Affect Each Other? Evidence from Student

Panel Data	1
1.1. Introduction.....	1
1.2. Background.....	6
1.2.1. Conceptual Framework.....	6
1.2.2. LEP Identification in the District	9
1.3. Empirical Strategy	11
1.4. Data.....	16
1.5. Results.....	19
1.5.1. Effect of Cohort LEP Share on Student Achievement	19
1.5.2. Effect of Cohort LEP Share on Mainstreaming of LEPs.....	21
1.5.3. Effect of Cohort LEP Share on Grade Retention.....	22
1.5.4. Heterogeneity in Effect of Cohort LEP Share by Student and School Characteristics.....	24
1.5.5. Investigation Mechanisms of the Effects of Cohort LEP Share	25
1.5.6. Instrumental Variable Estimates of the Effect of Exposure to LEP Students on Student Achievement of Ever-LEP Students	27
1.6. Conclusion	28

2. Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented

Programs on Students (with Scott A. Imberman and Steven G. Craig)	43
2.1. Introduction.....	43
2.2. The Gifted and Talented Program in LUSD	48
2.3. Model and Specification	50

2.3.1. GT Program Evaluation Using Regression Discontinuity.....	51
2.3.2. GT Magnet Evaluation Using School Lotteries	52
2.4. Regression Discontinuity Estimates of GT Impacts	54
2.4.1. Data.....	54
2.4.2. Tests of Validity of RD Design	54
2.4.3. Results	56
2.5. Estimates of the Impact of Attending a GT Magnet Using Randomized Lotteries	61
2.5.1. Data.....	62
2.5.2. Tests of Validity of Lottery Design.....	63
2.5.3. Results for GT Magnet Programs.....	64
2.6. Discussion	66
2.7. Summary and Conclusion	69
3. Are Remittances in the Hands of Women more Effective? Evidence from Vietnam (with Adriana Kugler)	87
3.1. Introduction.....	87
3.2. Related Literature.....	89
3.3. Data	92
3.4. Empirical Framework	94
3.5. Results.....	97
3.6. Conclusion	101

List of Figures

2.1. Gifted and Talented Matrix for GT Entry in 2008-09	71
2.2. Gifted Status in 7 th Grade by 5 th Grade Matrix Score	72
2.3. GT Qualification by Matrix Points	73
2.4. Gifted Status in 7 th Grade by Distance to Boundary Based on 5 th Grade Matrix Points	74
2.5. Distribution of Distances to Boundary	75
2.6. Reduced-form Effects on Achievement 7 th Grade by Distance to Boundary	76
2.7. Rank in Course by Final Grade in 7 th Grade by Distance to Boundary	77

List of Tables

1.1. Summary Statistics of the District’s Ever-LEP 5 th Graders from 1998-99 to 2009-10	30
1.2. Variation in Cohort LEP Share	31
1.3. Reduced-Form Effects of Cohort LEP Share on Student Achievement	32
1.4. Reduced-Form Effects of Cohort LEP Share on Mainstreaming	33
1.5. Reduced-Form Effects of Cohort LEP Share on Grade Retention	34
1.6. Heterogeneity in Effect of Cohort LEP Share by Student and School Characteristics.....	35
1.7. Regressions of Student Initial ESL/Bilingual Status on Cohort LEP Share	36
1.8. Examining the Effect of Cohort LEP Share on Student Achievement with Additional Controls	37
1.9. Relationship between Cohort LEP Share and Student English Proficiency	38
1.10. Effects of Exposure to LEP Students on Student Achievement of Ever-LEP Students	39
1.A1. Transition Matrix of Students Taking Aprenda Exam in 1 st Grade	40
1.A2. Shares of LEP Students Taking Stanford/Aprenda Achievement Exam	41
1.A3. Transition Matrix of LEP Students to Mainstreamed Classes	42
2.1. Characteristics of Students Evaluated for Middle School GT in 2007-08.....	78
2.2. Reduced-Form Estimates of Discontinuity in Pre-Existing (5 th Grade) Student Characteristics	79
2.3. Regression Discontinuity Estimates of Impact of Receiving GT Services.....	80
2.4. 2SLS Regression Discontinuity Estimates of Impact of Receiving GT Services Specification Checks	81
2.5. 2SLS Estimates of Impacts of GT Services Effects on Educational Environment and Student Choices	82

2.6. Balancing Tests for GT Magnet Lotteries – Covariates Measured in 5 th Grade.....	83
2.7. Effect of Attending a GT Magnet School Relative to a GT Neighborhood Program.....	84
2.8. Treatments from Attending a GT Magnet School Relative to a GT Neighborhood Program	85
2.9. 2SLS Estimates of Impacts of GT on Course Grades and Rank (2007-08 Evaluation Cohort).....	86
3.1. Basic Descriptive Statistics of Remittance-Receiving and Non-Receiving Households in the 1997-1998 Survey	102
3.2. First Stage Results.....	104
3.3. OLS and IV Estimates of the Impacts of Remittances on Education	105
3.4. OLS and IV Estimates of the Impact of Remittances on Health	106
3.5. OLS and IV Estimates of the Impact of Remittances on Employment	107
3.6. OLS and IV Estimates of the Impact of Remittances on Household Entrepreneurial Activities and Household Expenditures	108

Chapter 1

How do Limited English Proficient Students Affect Each Other's Educational Outcomes? Evidence from Student Panel Data¹

1.1. Introduction

Students are identified as having Limited English Proficiency if they do not speak English as their primary language and have limited ability to read, speak, or understand English.² The number of these students in U.S. public schools has increased substantially over time. From the 2000 U.S. Census, almost 20 percent of U.S. school-age persons speak another language at home.³ According to the National Clearinghouse for English Language Acquisition (NCELA), from the 1995-96 to 2005-06 academic years and for grades PK-12, English learners' enrollment increased by more than 57 percent, while overall enrollment increased by only 3 percent.⁴ As immigration continues to increase in the United States, addressing the academic needs of LEPs will be imperative.

Once students are identified as LEP, schools accommodate them with English as a Second Language (ESL) or Bilingual Education (BE) programs. In ESL programs, instruction is

¹ This project was conducted under the supervision of my advisors, Prof. Aimee Chin and Prof. Scott Imberman. I thank them. I also want to thank Prof. Steven Craig and the seminar participants in University of Houston graduate research workshop, the SEA, Prairie View A&M Research Seminar, and the AEFPP conference.

² Note that LEPs are also called English Language Learners (ELLs).

³ See <http://www.census.gov/population/cen2000/phc-t20/tab02.xls>.

⁴ See http://www.ncela.gwu.edu/files/uploads/4/GrowingLEP_0506.pdf.

done in English with supplemental language support. Students may have an in-class ESL teacher, or they may be placed into mainstream classes with occasional pull-out time to improve their English skills. In these mainstream classes, LEPs are taught alongside native speakers and other LEPs who speak the same or different home languages. In the anonymous urban school district from which I obtained my data, the ESL program serves mostly LEPs whose home languages are not Spanish, and when the number of LEPs in a language group in the district is not large enough to constitute a class. Another program is BE, in which students are generally placed in classrooms filled entirely with other LEPs who speak the same home language. In the anonymous school district, there are four types of bilingual education programs. Three are for students whose home language is Spanish.⁵ In these programs, instruction is done mostly in Spanish in the early grades. The amount of English instruction increases as students progress from grade to grade. The fourth type of BE was designed to serve LEPs whose home languages are non-Spanish.⁶ In this program, the students of the same home language receive primary language support for concept development and cultural enrichment activities. However, main language of instruction is English. In receiving these services, LEPs are exposed to peers who are also LEP.

Much of the discussion and research on the instruction of the LEPs revolves around whether it is better to teach them in their native languages or in English. Matsudaira (2005) exploits the quasi-random assignment of students to bilingual and mainstream (English immersion) classes to compare the effects of the two programs. Comparing the students' scores during the four years after the assignment, the author finds little difference between the two programs. The author stresses possible effects coming from peers, as BE students with low English ability and achievement are taught in classrooms with students from similar educational and cultural backgrounds. The author also notes that finding small differences between the two

⁵ Two of these programs serve only LEPs, while the goal of the third one is to develop full bilingualism for both Spanish speaking LEPs and English proficient students.

⁶ As of 2009-10, this program is only implemented for students from homes speaking Vietnamese. Students from Mandarin, Arabic and Urdu households will soon be served under this program.

programs suggests that peer quality does not matter in the context of immigrant children or that any positive effects coming from the differences in the two programs offset by possible (negative) effects from peer quality. Before this paper, meta-analyses have found conflicting results. While Baker and de Kanter (1981) and Rossell and Baker (1996) find little evidence to support BE, Greene (1998) finds evidence supporting BE over English immersion. Clearly, a consensus has not yet been reached in this literature and an important impediment is the paucity of studies with convincing research designs. Nonetheless, little is known about how interactions among LEPs are generated as the result of these programs, and the effects of such interactions on LEPs themselves. This chapter identifies the impact of exposure to other LEPs on the educational outcomes of LEPs themselves. It examines the existence and size of peer effects among LEPs.

Most economics studies of how peers influence each other in educational settings have focused on the effects of exposure to peers with higher and lower school performance on academic and behavioral outcomes of primary and secondary students,⁷ or on the effects of exposure to peers with more or less disruptive behavior.⁸ The focus of these studies has been the general student population. One of these papers, Betts and Zau (2004), looks at the impact of classroom and grade level peer achievement on individual elementary school students' rate of achievement. When Betts and Zau conduct the analysis on the subsample of English learners, they find that the classroom peer effect is highly significant, and is almost double the corresponding coefficient for all students.

Other papers have looked at the effects of exposure to peers from the same/different gender and racial group. For example, Hanushek, Kain, and Rivkin (2009) find that a higher percentage of black schoolmates reduces achievement levels for blacks. Hoxby (2000) finds that students are affected by the achievements of their peers, and intra-race peer effects are stronger. Lavy and Schlosser (2011) find that a higher proportion of females improves the cognitive

⁷ Hanushek, Kain, Markman, and Rivkin (2003); Angrist and Lang (2004); Hoxby and Weingarth (2006); Lavy, Paserman, and Schlosser (2008); Imberman, Kugler, and Sacerdote (forthcoming).

⁸ Figlio (2005); Aizer (2008); Carrell and Hoekstra (2010).

outcomes of both girls and boys. Black, Devereux, and Salvanes (2010) look at peer effects on students' long-run outcomes, such as labor outcomes and teenage births. They find that peer characteristics have little effect, except for the proportion of girls. For instance, they find that while a higher proportion of girls has a negative effect on the completed education of boys, it has a positive effect on the education of girls.

To my knowledge, two papers have looked at peer effects coming from students speaking a language other than English at home. First, Friesen and Krauth (2011) estimate the effect of the home language and other characteristics of a student's same-grade classmates on the achievement of that student. Using three cohorts of students in British Columbia, the authors estimate the effect of classroom composition on the achievement of students when they are in 4th grade, again when they are in 7th grade, and the students' achievement gains. They find that differences in peers' home languages affect student academic performance. Particularly, relative to English home language peers, there is a small net benefit coming from Chinese home-language peers on grade 4 scores and a cost from Punjabi home-language peers on grade 7 scores and score gain. While Friesen and Krauth (2011) focus on the effect of attending "enclave" schools on the achievement of students, my project focuses on the population of students who are most likely to be influenced by the spoken languages of their peers, the LEPs, and how being among LEPs affects their student outcomes. I estimate peer effects amongst these students with a much richer dataset and in a context that is more applicable to US education. Additionally, Friesen and Krauth's estimates do not distinguish between effects of exposure to a certain ethnic group from effects of exposure to less English proficient students, and likely many of the students with Chinese and Punjabi home languages in their dataset are proficient in English. Finally, besides examining the effect of the share of students who are LEP on student achievement, I also analyze its effect on the mainstreaming and grade retention of these students.

The second closely related paper studies how immigrant children affect the academic achievement of native Dutch children. Specifically, Ohinata and Ours (2011) use data from the

2001 and 2006 Progress in International Reading Literacy Study (PIRLS) and the 1995 and 2007 Trends in International Mathematics and Science Study (TIMSS) to analyze the effect of concentration of immigrant children on the achievement of native Dutch children. To overcome the endogeneity problem of school attendance, the authors control for school fixed effects. They do not find strong evidence of negative spillover effects on the test scores from immigrant children to native Dutch children. However, they find that immigrant children themselves experience negative language-related spillover effects from having more immigrant children in the classroom but no spillover effects on the math and science scores. While Ohinata and Ours (2011) focus on the effect coming from immigrant children, my chapter concentrates on the effects coming from those with limited English proficiency to get at the language spillover. In general, peer effects from LEPs are not the same as peer effects from immigrants. According to the Urban Institute, 40 percent of LEPs, ages 5-19 are foreign born.⁹

Empirical identification of the effect of peers who are LEP on other LEPs is difficult because LEPs are not distributed randomly across schools. Families select where they live and where to send their children to school, while school administrators assign students to a particular class for special instruction or instruction from a specific teacher. As the result, estimates of the effect of the share of LEPs in a grade on LEPs will capture both peer effects and effects coming from the unobserved characteristics that also affect the outcome variables and peer selection. To deal with the selection bias, I utilize a rich panel data set and employ various fixed effects. By focusing on specifications that incorporate school fixed effects and school-specific time trends, I am able to control for unobservable characteristics that might correlate with the LEP share. This empirical strategy depends on there being some variation across cohorts' LEP shares within a

⁹ See http://www.urban.org/uploadedpdf/410654_NABEPresentation.pdf.

school that is idiosyncratic.¹⁰ Such cohort-to-cohort variation might arise from natural population fluctuations, including from births and migration.

Estimating the reduced-form effects of cohort LEP share on the educational outcomes of LEPs, I find the share of LEPs in a cohort has positive and significant effects on academic achievement of LEP 5th graders, particularly in math scores. These results appear to be more pronounced for males and for students coming from schools with more low socio-economic status students. I also find significantly faster mainstreaming of LEPs. The by-grade analysis suggests that students are most likely to be mainstreamed at the first opportunity, which is the year after first grade. Finally, I find that students with higher cohort LEP share are less likely to be retained. The negative effect of cohort LEP share on retention appears to be larger for students from schools with more low socio-economic status students and for students with lower initial achievement.

These reduced-form effects can be the result of many mechanisms, and I investigate a few of these empirically. I provide evidence that the positive LEP peer effects do not come from student differential enrollment in BE and ESL programs, differential peer achievement, or differential peer English-language proficiency. The remainder of the chapter is organized as follows. Section 1.2 discusses the background information, specifically about the conceptual framework and LEP identification at the school district. Section 1.3 explains the empirical strategy. Section 1.4 describes the data. Section 1.5 presents the results. Section 1.6 concludes.

1.2. Background

1.2.1. Conceptual Framework

The share of LEPs may affect the academic outcomes of LEPs in various ways. One possible channel is by changing the speed of English language acquisition. There are many

¹⁰ Hoxby (2000) is one of the earliest to apply this strategy, and she uses it to identify racial and gender peer effects on student achievement.

studies that have linked inadequate English proficiency to low student achievement (Clewell, Fix, and Ruiz-de-Velasco, 2000; Suarez-Orozco and Suarez-Orozco, 2001; August and Shanahan, 2006). When LEPs and proficient speakers of English are given the same math exam written in English, the former group scores lower than the latter. LEPs perform better when they receive linguistic accommodation, such as modification of math questions or simplified English dictionaries during exams (Abedi and Lord, 2001; Albus, Thurlow, Liu, and Bielinski, 2005). Having higher English proficiency increases understanding of the math questions and appears to improve English reading scores. Also, the sooner the LEPs become proficient in English, the earlier the students spend all their time in learning the content of the class, instead of working on improving their English skills. Moreover, having low English proficiency may act as a signal of low student achievement. As a result, LEPs may not be selected to take more challenging courses, lowering their achievement (Wang and Goldschmidt, 1999).

Being surrounded by other students who are also LEPs may affect the English acquisition of LEPs. The amount of time immigrant adolescent students spend speaking English in informal settings is predictive of English language proficiency (Carhill, Suarez-Orozco, and Paez, 2008). Being among students with limited English proficiency may reduce the quantity and quality of LEPs' conversations in English. For one thing, students may utilize their home languages more in conversations when speaking to classmates who are from the same home language group (Willoughby, 2009). In speaking to other LEPs whose home languages are different, LEPs use English, but due to the students' limitations in their English proficiency, they expose each other to more broken English.

Additionally, studies in the area of English fluency suggest the size of one's language group affects one's English proficiency. Lazear (1999) theorizes and finds evidence in the 1900 and 1990 US censuses that immigrants from groups with large proportions in the local population learn English more slowly than immigrants from groups with smaller proportions in the local population. Angrist, Chin, and Godoy (2008) find that the 1949 change of language of instruction

in public schools from English to Spanish in Puerto Rico had little effect on Puerto Rican English-speaking ability. This suggests that language environments are relevant in the process of learning English. For LEPs, the language they hear from their classmates matters, not just the language of instruction. For instance, a large group of students speaking the same home language may have less incentive to learn English, as the students can rely on others to help them to understand the classroom materials.

LEPs may affect each other's academic outcomes through other channels, besides that of English learning. Among all students, including LEPs, peer effects involve students teaching one another. Even though LEPs may have similar level of English proficiency, those who are immigrants enter American schools with different academic histories (Suarez-Orozco and Suarez-Orozco, 2001). Thus, within a classroom, there is diversity in terms of student knowledge of the classroom materials. Additionally, a student's innate ability and knowledge of his/her home culture and language can affect his/her peers by ways of knowledge spillover and through adjustments to the classroom standard of English language proficiency. Coming from culturally diverse homes, LEPs may influence their peers through different attitudes toward schooling and habits of study. Not only do students coming from different families affect other students, but students from different cultures also affect each other.

Furthermore, the effects of LEP share on the outcomes of LEPs can come from classroom instruction. The mechanisms are similar to those of ability grouping. Teachers have students who are similar in English proficiency. As a result, the teachers can focus their instruction of these students. The students receive specialized instruction appropriate to their level of English proficiency, including instruction that furthers their English acquisition. Additionally, due to the prevalence of LEPs throughout the school, teachers in regular classes may be more familiar with the best practices of instructing LEPs, and will know how to accommodate them if there are one or two LEPs in their classrooms.

To summarize, there are various channels through which cohort LEP share can affect student achievement. While interactions among LEPs may result in slower English acquisition, which reduces student educational outcomes, these interactions may result in student learning from each other, which might have a positive effect on student outcomes. While being placed in the same classroom means that LEPs are less exposed to native speakers, this grouping allows teachers to deliver a specialized curriculum to LEPs, which might translate into more effective learning. Together, the effect on student outcomes can be either negative or positive, depending on the size of each effect. My reduced-form estimates will encompass all these various effects, but I will be able to speak to the roles of a few potential mechanisms. In particular, I will consider whether impacts work through achievement peer effects or changes in instructional settings.

1.2.2. LEP Identification in the District

In the district I study, state law delineates the identification of LEPs. A committee set up at each school is in charge of this process and of the placement of LEPs in the appropriate program available at each school. The identification process begins with the Home Language Survey, which is sent home to the parents to be completed within four weeks of each student's enrollment to determine the spoken language in the student's home and the language that the student uses at home. If the answer to either question is a language other than English, then the student's language proficiency is evaluated through oral exams and/or the student's past or current written reading and language tests. If a new student is identified as LEP, the committee must determine his/her English proficiency level and the most appropriate instructional program for him/her.

Parents' approval for their children's enrollment in a BE or ESL program must be obtained. If the parent wishes to waive program participation, the LEP student is assigned to all

English classes, and receives regular instructional and testing programs.¹¹ However, this student is considered LEP until he/she meets the exit criteria.

All students from grades 1-11 in the district take nationally norm referenced achievement tests. Non-Spanish speaking LEPs take the Stanford Achievement Test, while Spanish speaking LEPs take the Aprenda exam, which is an exam written in Spanish and modeled after the Stanford exam. Examining all LEP students in the district from 1st to 5th grade from school years 1998-99 to 2009-10 reveals that in 1st grade the percentages taking Stanford and Aprenda were 20 and 75, respectively. In grade 5, the two numbers were 83 and 13, respectively. It appears that I do not have data on achievement for all students in every grade.¹² One of the potential reasons for missing achievement scores is that the committee may exempt Year 1 immigrant LEPs who are determined to be non-literate at time of entry.¹³ The school committee determines which test each LEP student takes based on the language used most for reading/language Arts instruction. Students also take an English language proficiency exam. This exam examines the students' progress in learning the English language in four language domains: listening, speaking, reading, and writing. Students' scores come in the forms of ratings from 1 to 4, which translate to beginning, intermediate, advanced, and advanced high, respectively.

All LEPs must be formally exited from the program. LEPs in PK-1st grade may not be exited from the BE or ESL programs. If the student meets the exit criteria, he/she will be assigned regular classes at the beginning of 2nd grade. The criteria to exit from LEP services include being classified as Fluent or as English Speaker in the IDEA Proficiency Test (a test of oral language proficiency) for all grades, and scoring greater than the 40th percentile in both Stanford reading and language tests in grade 1-2, or scoring at least "passing" in the reading and ELA (English language arts) English sections of the state standardized exam in grades 3-12. To tabulate some

¹¹ From the data, about 6 percent of the observations with non-missing information on parents' approval have parents waiving LEP program participation.

¹² Having missing achievement scores does not appear to have any significant relationship with the variable of interest.

¹³ See Table 1.A2 in Appendix for more details.

statistics regarding mainstreaming at the district, I construct a sample of data limited to observations of students who are LEP in 1st grade, and I observe them in grades 2-5 for the school years 1998-99 to 2009-10. In grade 2, about 8 percent of the students are mainstreamed. By grade 5, 40 percent of the students are mainstreamed.¹⁴

1.3. Empirical Strategy

Before describing my empirical strategy, a unique issue about examining the achievement test scores of LEP students should be mentioned. Spanish speaking LEPs are assessed in Spanish and others in English. It is likely endogenous what language the exams are taken by LEPs in this school district. Higher ability students may be given opportunity to take the Stanford exam earlier and may be mainstreamed faster, and higher quality schools may mainstream their students faster as well. Additionally, Akresh and Akresh (2011) find that depending on the students' English proficiency, students' scores on an exam vary with the language of the exam. Looking at statistics, the majority of students who take the Aprenda exam in 1st grade continue to take Aprenda in 2nd and 3rd grade. However, by 5th grade, less than 1 percent of these students take the Aprenda exam.¹⁵ Given these statistics, to limit the bias coming from the endogeneity in the type of exam the students take, I restrict my analysis of test score outcomes to 5th graders. Another reason for looking at the 5th graders is that I am able to estimate a longer run effect of exposure to LEPs, so my estimates reflect accumulated effects of exposure to students with limited English proficiency.

A major empirical challenge in estimating peer effects is the potential correlation between the share of peers who are LEP and a number of factors that are unobservable to the researcher, such as parental preference of residential location and school, neighborhood

¹⁴ See Table 1.A3 in Appendix for more details.

¹⁵ See Table 1.A1 in Appendix. This is also consistent with the practice at the district, whereby qualified LEPs enter a pre-exit phase in 4th or 5th grade. Here students' language of instruction is mostly English, and student assessments change from the Aprenda to the Stanford exam.

characteristics, and school quality. These unobservable factors generally affect both the share of peers who are LEP and the LEPs' own achievements and behavior. This suggests that traditional estimates of the share of LEPs on the student outcomes of LEPs may be biased.

I rely on variations across cohorts within schools by including school fixed effects to identify the effect of LEP share on the outcomes of LEPs. Particularly, I estimate a regression model similar to

$$Y_{ics} = \alpha + \beta AVGLEPShare_{ics} + \Omega \mathbf{X}_{ics} + \Pi \mathbf{C}_{cs} + \Phi \mathbf{G}_{cs} + \eta_s + \varepsilon_{ics} \quad (1)$$

on 5th grade LEPs, where the unit of observation is LEP 5th grader i in cohort c and school s . Variable Y is an educational outcome of interest of LEP 5th graders; $AVGLEPShare$ is “average grade LEP share”, which is the share of students in a grade classified as LEP averaged over the grades the 5th grade LEP student has been through in the district. This variable provides a measure for an average exposure to LEPs. Matrix \mathbf{X} includes student characteristics, namely gender, ethnicity, initial achievement, and indicators of whether the student is economically disadvantaged, a recent immigrant, a gifted student, or a special education student. The term “initial” means 1st grade or the first grade the student was in the district. Student initial achievement consists of student's initial scores in math, reading, and language, three dummy variables indicating whether the respective exam was Stanford (or Aprenda), and interactions of the score and its corresponding Stanford exam indicator.¹⁶ It also contains year fixed effects, which control for the common time effects across all schools. Matrix \mathbf{C} includes shares of student characteristics in the 5th grader i 's 1st grade cohort. These shares are included in the regression to capture the characteristics of the group of students to whom the 5th grader i is supposed to have exposure to throughout his elementary school had everyone in his cohort entered the district at the same time, assuming normal grade progression. \mathbf{G} includes shares of student characteristics in the students' 5th grade cohort. These shares of student characteristics are

¹⁶ I do not have enough data on the student English proficiency exam to control for student initial English proficiency.

the share of students who are economically disadvantaged, the share of students who are female, the share of students who are gifted, in special education, Native American, Asian, Black, and White. These shares are included to capture the observable characteristics of the student i 's 5th grade cohort. η_s is a school fixed effect that absorbs permanent unobserved factors of the school, such as school quality, which may affect both LEP share and student achievement.¹⁷ In this context, school fixed effects are meaningful controls for selection across schools. For example, schools may have unobservable characteristics that attract LEPs, and these factors also affect student outcomes. With the school fixed effects, I am able to control for these unobservable characteristics such as a school's general educational environment and quality, which are likely to be constant over time. The term ε_{ics} is the random error, which might be correlated across student observations from the same school and across time, and might include an individual random element.¹⁸

One concern is that parents and school administrators may have discretion in placing LEPs into a particular classroom. To avoid the selection at the classroom level I focus on analyzing the effect of average LEP shares in the grade rather than shares in the student's classroom on outcomes of LEPs.

The second concern is that over the 12 years of the panel, there may be time-varying unobservable factors that are also correlated with the LEP shares and with student outcomes within the same school. Specifically, these time-varying unobservable factors such as teacher quality may be trending over time in a way that is correlated with LEP share. For example, schools successful at teaching LEPs may attract more LEPs over time. As a result, schools might show systematic trends in the share of their LEPs. Thus, unobserved differences in students across cohorts, and within the same school, could be correlated with differences in the LEP shares

¹⁷ By including student fixed effects, the variation in a student's LEP share would come from grade promoters, repeaters, students switching schools, and new immigrants, hence I exclude student fixed effects.

¹⁸ I use robust standard errors that are clustered at the school level (Bertrand, Duflo, and Mullainathan (2004)).

and confound estimates of the effect of LEP shares. To address this concern, I control for school specific linear trends by running the following regression model:

$$Y_{ics} = \alpha + \beta AVGLEPShare_{ics} + \Omega X_{ics} + \Pi C_{cs} + \Phi G_{cs} + \rho_s \times cohort_c + \eta_s + \varepsilon_{ics}(2).$$

Equation (2) is similar to that of (1), except for the school-specific linear time trends factor, $\rho_s \times cohort_c$. Here, identification of average grade LEP share comes from the deviation of the average grade LEP share from its long-term trend within a school.

Finally, the third concern is that the LEP shares may be endogenous, as students get evaluated for mainstreaming every year. This happens because LEP statuses can change. When students enter the district, schools evaluate their English proficiency and identify their LEP statuses. However, by 2nd grade, LEP students can be mainstreamed. Once the students are mainstreamed, they are no longer identified as LEP. Thus, LEP status correlates with unobservable characteristics such as school, teacher, and LEP program quality. Accordingly, I focus my analysis on students who have ever been classified as LEP¹⁹. I call these students ever-LEPs. Additionally, the average grade LEP share or *AVGLEPShare* comes from the share of students who are LEP in a given grade and school. By definition, this share is the same as the 1st grade cohort LEP share for a given cohort if students do not repeat or jump a grade, no student leaves/enters the school during the time frame, and no student gets mainstreamed. In the event the LEPs get classified as non-LEP, the average grade LEP share will be smaller than the 1st grade cohort LEP share as students progress from grade to grade. More worrisome is the possibility that the differences between the two variables will be correlated with the unobservable characteristics that also affect student outcomes. For example, students may be mainstreamed faster at a higher-quality school. If the mechanisms that dictate the higher rate of mainstreaming also affect achievement and retention, then my estimates of the average grade LEP share will be biased. To address the possibility that *AVGLEPShare* is endogenous, I measure exposure to LEPs using 1st

¹⁹ This definition retains students' initial LEP statuses if they enter the district before grade 1, as the general practice is that if a student is LEP prior to grade 1, he/she can only be mainstreamed at the beginning of grade 2.

grade cohort LEP share, which is the share of students in 5th grader i 's 1st grade cohort who are LEP. This measure captures initial exposure to LEPs, before any mainstreaming which is potentially endogenous as explained above. In Section 1.5.6, I use 2SLS to estimate the effect of average grade LEP share on outcomes of the ever-LEPs using 1st grade cohort LEP share as the identifying instrument. For now, my interest lies in estimating the reduced-form effects of cohort LEP share on student outcomes. Specifically, I estimate reduced-form effects of cohort LEP share on student outcomes following the econometric model below:

$$Y_{ics} = \alpha + \beta cohort_LEP_share_{cs} + \Omega X_{ics} + \Pi C_{cs} + \Phi G_{cs} + \rho_s \times cohort_c + \eta_s + \varepsilon_{ics} \quad (3),$$

where $cohort_LEP_share_{cs}$ is the share of students who are LEP in the student's 1st grade cohort, and cohort LEP share henceforth.

Essentially, the panel data model with school fixed-effects and school-specific time trends is relying on cohort variations within schools to identify the effect. There are works that rely on idiosyncratic variations in adjacent cohorts for identification of peer effects. Hoxby (2000), Lavy and Schlosser (2011), and Black, Devereux, and Salvanes (2010) use idiosyncratic changes in gender composition to look for gender peer effects. Gould, Lavy, and Paserman (2009) use idiosyncratic variations in the proportion of immigrants to study the impact of immigrants on student performance. Hoxby (2000) also uses the cohort-to-cohort changes in the share of students who are of a particular racial ethnic group to find racial peer effects. My methodology is similar to these papers. The basic idea is to compare outcomes of LEPs from different cohorts who have similar characteristics and face the same school environment, except for the fact that the share of LEPs is higher in one cohort than another due to random factors. Similar to these papers, my key identifying assumption is that conditional on student, grade and cohort observable characteristics, school fixed effects, and school specific trends, the cohort LEP shares are uncorrelated with changes in unobservable factors that could affect students' educational outcomes. This assumption is reasonable as I rely on the natural fluctuation in the

number of children who speak a language other than English at home. For example, in this school year, 30 percent of the 1st graders are LEP. In the following school year, that number is likely to be different, as the number of total students and the number of LEPs in a particular neighborhood depends on who lives in the neighborhood and the age distribution of the school-aged children. I rely on these small changes in cohort LEP share to identify the effects of cohort LEP share on student outcomes. Parents may know the average LEP share at a school, but the parents may not be able to anticipate the changes in the LEP shares from one cohort to another, and it would be much harder for parents to anticipate the deviations from school trends in LEP share.

1.4. Data

My data consist of administrative records for students from a large urban school district in the Southwest United States. The data set comprises of student demographics, including gender, ethnicity, economic status, gifted status, special education status, recent immigrant status, home language, and test scores for every student in the district up to 12th grade. Testing data include the Stanford and Aprenda achievement test scaled scores. Separately, for each exam I standardize the scaled scores across the district within grade and year using all students in the district. Testing data also include the English Language Proficiency exam ratings. However, while other data are available from the 1998-99 to 2009-10 school years, these ratings are only available from school year 2004-05 to school year 2009-10. In addition to the ratings on students' language domains of listening, speaking, reading, and writing, I have a composite rating. This composite rating is determined from the four language domains, with the highest weight of 75 percent being placed on the reading rating.

To obtain the sample used for my empirical analysis, I first restrict the data to grades 1-5. I define the grade the student enters into the district as the earliest grade he/she appears in the data. I also define a student's LEP status as the LEP status when he/she first entered the data. This is the student's ever-LEP status. I construct the LEP share as the share of students who are LEP in

a grade and school. Next, I assign each student a 1st grade cohort. A student's cohort is a group of students who start a grade at the same school and year. For students who enter the district after 1st grade, which is less than 25 percent of the sample, I assign them to the 1st grade cohort to which they would have belonged had they been in their entering school since 1st grade assuming normal grade progression. For all the student characteristics including economically disadvantaged, gifted, and special education status, sex and ethnicity, I also construct share of student characteristics in a grade and school, and in a cohort and school in similar fashion.

Lastly, to obtain a cumulative measure of 5th graders' exposure to LEPs, I construct a measure that encompasses their exposure to LEPs over the length of time they were in the district. This variable is the "average grade LEP share", i.e. *AVGLEPShare*. The variable gives the share of students in a grade classified as LEP averaged over the grades the student has been in the district.

My panel data set spans from the 1998-99 to 2009-10 school years. I drop students who are never classified as LEP, observations coming from students' first-year in the sample, and observations of students who have not yet reached grade 5. For those who repeat 5th grade, I keep only their first 5th grade observations. Additionally, after dropping observations with missing data on variables used in the empirical analysis, I have a total of 44,436 ever-LEP students in 216 schools. Within this sample, almost 84 percent of the students observed entered the district in 1st grade or earlier.

My achievement outcomes are the Stanford scores in math, reading, and language. In addition to achievement, from a policy standpoint, mainstreaming and retention are important educational outcomes for LEPs. A student is mainstreamed if the student meets the LEP exit criteria, and thus is no longer classified as LEP. These mainstreamed students take regular classes. To assess the effect of exposure to LEPs on student grade retention, I construct a dummy variable indicating whether the student was ever-retained since 1st grade or since he/she entered the school district, if the student entered the district after 1st grade.

Panel A of Table 1.1 provides summary statistics for all ever-LEP 5th graders in the district. Regarding achievement scores, while LEPs' average scores on math gather around the mean of zero, scores for reading and language are much lower, 0.22 and 0.17 of a standard deviation below the math score, respectively. On average, cohort LEP share is 56 percent. The average "average grade LEP share" is about 50 percent. About 95 percent of the students are economically disadvantaged by definition of being eligible for free or reduced price lunches or being otherwise economically disadvantaged. Almost 95 percent of the students are Hispanic. More than 62 percent of the students are classified as LEP.

Next, I divide the full sample into two subsamples, depending on the observation's cohort LEP share. Panel B is restricted to observations of those with cohort LEP shares above the full sample's median of cohort LEP share, and Panel C is restricted to observations of those with cohort LEP shares equal to or below that figure. Simple comparisons between the statistics in Panels B and C reveal that students with higher cohort LEP shares tend to have lower 5th grade achievements relative to students with lower cohort LEP shares. They also appear to be more likely to be recent immigrants and Hispanics. Their schools also seem to be larger, with more students enrolling in BE and fewer students enrolling in ESL classes.

The dissimilar statistics of the high cohort LEP share and lower cohort LEP share subsamples suggest that a simple OLS regression of exposure to LEPs on student outcome will be biased. Even if I control for their observable characteristics there are likely many unobservable underlying factors that affect both the measures of exposure to LEPs and student outcome at the school, providing biased estimates. As a result, I apply an empirical strategy that utilizes idiosyncratic variation of cohort LEP shares within schools to estimate the effect of exposure to LEPs on student outcomes.

1.5. Results

My key identifying assumption is that conditional on student, grade, and cohort observable characteristics, school fixed effects, and school specific trends, the cohort LEP shares are uncorrelated with changes in unobservable factors that could affect students' educational outcomes. Following Lavy and Schlosser (2011) and Bifulco, Fletcher, and Ross (forthcoming), I first examine whether there is sufficient variation in cohort LEP shares after controlling for school fixed effects and trends to obtain precise estimates.²⁰ Table 1.2 displays the variation of cohort LEP shares as raw data and the variation that is left after removing school fixed effects and trends. Standard deviations are reduced by almost 70 percent after removing school fixed effects and trends. This means that most of the variation is across schools as it disappears when school fixed effects are included. I rely on the remaining variation to estimate the impact of cohort LEP share on student outcomes. Table 1.2 suggests I have sufficient variation to estimate the effects. The variation in cohort LEP share after the removal of school fixed effects and trends is more than double the variation in student composition within schools reported by Bifulco, Fletcher, and Ross (forthcoming). That amount of variation was enough for the authors to obtain significant estimates of the effect of cohort composition on outcomes of high school students.

1.5.1. Effect of Cohort LEP Share on Student Achievement

Table 1.3 reports reduced-form estimates of the effects of cohort LEP share on student achievement using the sample of all ever-LEP 5th graders who have been in the district for at least a year. Outcome variables are student 5th grade standardized Stanford achievement scores. Their means and standard deviations are listed in the table.

Column 1 lists the coefficients of cohort LEP share from regressions of student achievement on cohort LEP share controlling for student characteristics and initial achievement,

²⁰ I do not provide results of a “balancing test” as in Lavy and Schlosser (2011) and Bifulco, Fletcher, and Ross (forthcoming), because by construction cohort LEP share correlates with many of the student characteristics in my data.

as well as year fixed effects. Student characteristics are gender, economically disadvantaged status, recent immigrant status, gifted status, special education status, and student ethnicity. The estimated effects of cohort LEP share on student achievement are negative. They are large and significant for reading and language scores. This relationship is consistent with the descriptive statistics in Table 1.1, where schools with higher cohort LEP share tend to have lower student achievement.

Some cohort-grade characteristics are added to the previous specification and the results are reported in column 2. These variables are shares of student characteristics, except for recent immigrant status, in the student 1st grade and 5th grade cohorts.²¹ With the addition of cohort controls, the estimates are no longer significant, but are still negative. However, when I add campus fixed effects in the third specification, the math and reading estimates turn positive and remain insignificant as shown in column 3. This change in sign is a result of using the variation across cohorts within schools and throwing out cross-school variation to identify the effects of cohort LEP share. Next, in column 4, school specific time trends are added as controls. This specification is my preferred specification, as it controls for possible school trends in addition to campus fixed effects, and it identifies the effects using the deviations from school specific time trends in cohort LEP share.

Only the effects of cohort LEP share on math scores are significant at the 5 percent level. Specifically, a one standard deviation or a 20-percentage point increase in cohort LEP share increases the math scores in 5th grade by more than 0.02 of a standard deviation. For reading and language scores, the estimates are positive but not statistically significant. Hence, overall there appears to be no negative effect of cohort LEP share on reading and language scores, and there is evidence of positive effect on math.

²¹ The shares of students who are recent immigrants in a cohort and a grade are excluded to lessen the concern of co-linearity with the variable of interest, as a recent immigrant student is likely to be limited in English proficiency.

1.5.2. Effect of Cohort LEP Share on Mainstreaming of LEPs

Once LEPs gain enough proficiency in English to function in mainstream classes (as determined by a committee - see Section 1.2.2), then they exit LEP status and cease to receive BE or ESL instruction. Mainstreaming is a goal of educational programs for LEP students. In Table 1.4, I analyze how cohort LEP share affects the mainstreaming of LEP students, I first examine how cohort LEP share affects the number of years the ever-LEP students were classified as LEP in elementary school using the full sample of 5th graders who have been in the district at least a year and who are 1st grade or prior entrants.²² Panel A of Table 1.4 reports results of this analysis. On average, ever-LEP students are classified as LEP for about 4.3 years out of potential five years (grade 1-5) in elementary school.²³ The simple OLS regression with initial achievement, student characteristics, and year fixed effects as controls reveals little relationship between cohort LEP share and the number of years ever-LEP students were classified as LEP as reported in column 1. However, starting with the next specification, when controls are added, we begin to see a pattern. The estimates turn negative, get larger in magnitude, and become significant as controls and fixed-effects are added. Under the preferred specification, we see that higher cohort LEP share significantly reduces the number of years ever-LEP students are classified as LEP in elementary school. Specifically, a one standard deviation or a 20-percentage point increase in cohort LEP share decreases this time by 0.064 years or 0.57 months on a 9-month school year.

I find that ever-LEP students are mainstreamed faster. When is this faster mainstreaming occurring during students' academic career? In panel B of Table 1.4, I examine how cohort LEP share affects the probability of LEPs being mainstreamed by a given grade.²⁴ In this analysis, I limit observations to a sample of students who were LEP in 1st grade. I follow these students for the next four years. My outcome variable is a dummy variable that is equal to 1 when the student is no longer classified as LEP and thereafter. Thus, the coefficient for cohort LEP share gives the

²² Limiting the sample to 1st grade entrants dropped 7,170 observations from the sample.

²³ The measure was obtained after limiting the original sample of data to grades 1-5.

²⁴ Students are mainstreamed when they meet exit criteria as described in Section 1.2.

effect of cohort LEP share on the cumulative rate of mainstreaming up to the given year. The results suggest that the students are significantly likely to be mainstreamed in the 2nd year, which is the earliest LEPs can be mainstreamed. Looking at year 2 results, under the preferred specification, a one standard deviation or a 20-percentage point increase in cohort LEP share increases the probability of LEP students being mainstreamed by about 2 percentage points. The effect size is about 20 percent of the mean of being mainstreamed in year 2, which is 0.092. For years 3-5, cohort LEP share does not have a significant effect on mainstreaming, suggesting that students with lower cohort LEP share caught up with students with higher cohort LEP share. In other words, higher cohort LEP share causes some LEP 1st graders to be mainstreamed sooner – in second grade rather than third grade, and that there is little differential mainstreaming by cohort LEP share in later grades of elementary school.

There are a few possible interpretations of this result. One is that having more LEPs drives schools to do a better job at monitoring and evaluating them for mainstreaming. Second, having more LEPs may enable schools to focus their instruction of the LEPs more effectively and as a result students exit LEP services faster. Third is that schools are simply overwhelmed with large numbers of LEPs, and thus due to limited resources, they mainstream LEPs faster. The first two items are likely to positively affect student achievement. The last item is likely to negatively affects it. Since there is evidence of improvement in math, so on net it seems that the positive effect outweighs the negative effect.

1.5.3. Effect of Cohort LEP Share on Grade Retention

Table 1.5 reports the reduced-form estimates of the effect of cohort LEP share on grade retention. I examine the effect on the probability of a student ever being retained since 1st grade or since their entrance into the district if the grade of entrance is after 1st grade. The table reports coefficients of cohort LEP share on the regression of a dummy variable that equals 1 if the ever-LEP student was ever retained on the cohort LEP share. Under the preferred specification, a one

standard deviation or a 20-percentage point increase in cohort LEP share reduces the chances of ever-LEPs ever being retained by 2.5 percentage points for the sample limiting to 1st grade entrants though the effects are smaller for the full sample. It appears that the 1st grade entrants drive the effects on grade retention. One possible reason for this result may be that the restricted sample has less measurement error. The full sample consists of late entrants whose values on ever retained may contain measurement error since for the late entrants their pre-entry retentions are unknown.

Overall within a school, students with a higher cohort LEP share are less likely to be retained. This finding is economically important, especially when considering the evidence suggesting retention is associated with likelihood of dropping out (Rumberger, 1987; Grissom, and Shepard, 1989; Fine, 1991; Roderick, 1994). Jacob and Lefgren (2009) find that retaining low-achieving 8th grade students in elementary school increases the probability that these students will drop out in high school. The finding is more substantial for the school district I study, as 95 percent of my sample is Hispanic. The National Center for Education Statistics reports that the high school dropout rate among Hispanics was 17 percent in 2009, and that Hispanic high school dropout rate is the highest among all five ethnic groups.²⁵

This negative effect on grade retention may be a result of constrained resources. Due to having more LEPs, schools may not have enough resources to retain students to ensure maximum learning. Or maybe having more LEPs enables schools to focus their instruction of the LEPs more effectively. The first possibility is negative. The second is positive. Since there is evidence that math achievement increases, so it appears that the positive effect outweighs the negative effect.

²⁵ See <http://nces.ed.gov/fastfacts/display.asp?id=16>.

1.5.4. Heterogeneity in Effect of Cohort LEP Share by Student and School

Characteristics

Next, I examine heterogeneity in the effect of cohort LEP share by student characteristics. To do this, I divide the samples into subsamples and analyze the effect of cohort LEP share on student outcomes separately. I divide my samples into two subsamples according to five different characteristics: (1) whether a student is female or male, (2) whether a student is a recent immigrant or not, (3) whether the school's mean economic disadvantage is above the sample's median or not, (4) whether the school's mean enrollment is above the sample's median enrollment, and (5) whether the student's composite initial achievement is above the sample's median composite initial achievement or not. A student composite initial achievement was calculated by taking an average of the student's standardized scores in math, reading and language. When looking at the heterogeneity of the effect of cohort LEP share on student achievement I use the full sample of all ever-LEP 5th graders who have been in the district at least a year. With the mainstreaming and grade retention outcomes, the sample is restricted to 1st grade entrants.

The estimates reported in Table 1.6 come from the preferred specification, which includes controls, school fixed effects, and school specific trends. The effects of cohort LEP share on student achievement are insignificant for reading and language scores for all subsamples, except for the subsample of schools having mean enrollment greater than the sample's median enrollment. However, it appears that the significant effects of cohort LEP share on math scores are more pronounced for males, recent immigrants, students from poorer schools, students in larger schools, and students with higher initial composite achievement.

With the mainstreaming and retention outcomes, there appear to be three noticeable differences in the effect of cohort LEP share. One is that the effects on mainstreaming and retention are higher for schools with more low socio-economic status students. Secondly, the

effect of cohort LEP share on retention is higher for students from smaller schools. Thirdly, the effect of cohort LEP share on retention is larger for students with lower initial achievement. The findings suggest two things. One is that the poorer schools appear to drive the effect of cohort LEP share on math scores, mainstreaming, and retention. Secondly, while the effect on mainstreaming differs by initial achievements, the effect on retention is much higher for students with lower initial achievement, who are likely to be the marginal students.²⁶

1.5.5. Investigating Mechanisms of the Effects of Cohort LEP Share

I find strong evidence of cohort LEP share on ever-LEP students' educational outcomes. What are the specific mechanisms underlying these effects? Section 1.2.1 discussed some potential mechanisms and in this section I empirically investigate a few of them. One possibility is that my estimates of peer effects may pick up effects from BE/ESL programs. After all, it is the practice in the district to assign students to either the BE or the ESL program upon LEP identification. Thus, cohorts having higher cohort LEP shares may experience expansion to or improvement of BE and ESL programs, or differences in how students are assigned to the programs. To test this possibility, I run regressions of students' initial BE/ESL statuses on cohort LEP share following the 4 specifications as in Table 1.3. These results are provided in Table 1.7. With student initial BE status, OLS reveals a positive correlation: higher cohort LEP share is positively associated with a higher probability of the student being in the BE program when he/she enters the data. However, this selection story disappears in subsequent specifications. Within-campus regressions suggest there is no significant relationship between cohort LEP share and students' placement into BE/ESL program. In conclusion, there is little evidence to suggest that cohort LEP share affects student outcomes through differential assignment to BE and ESL programs.

²⁶ When I run fully interacted models, I find the only statistically differences are the effects of cohort LEP share on retention for recent immigrants and others and for students whose composite initial achievement are above or below the sample's median.

A second possibility is that the effects of cohort LEP share reflect peer achievement effects. Among LEPs there may be variations in achievement such that students' educational outcomes can be impacted by their peers' achievements. To test whether my findings on the effects of cohort LEP share pick up effects from peer achievement, I run a regression of the preferred specification with controls, school fixed effects, and school specific trends with the inclusion of initial peer achievement. Controls for initial peer achievement enter the regression through two variables: Initial peer achievement, which is the average of achievement in math, reading, and language of the student's peers, and the share of peers who initially take the Stanford (English) math, reading, and language exam. Peers are students in the same grade minus the student. Table 1.8 reports the coefficients of cohort LEP share and initial peer achievements in column 2. Column 1 reports the findings from column 4 of Table 1.3. When controlling for peer initial achievements, coefficients on cohort LEP share are similar to the ones in the left column. Thus, there is little evidence to suggest that the effects of cohort LEP share come from peer achievement

Lastly, one of the channels discussed above was the channel of English language acquisition. Using available data on the ratings of students' English proficiency, I present results from OLS regressions of 5 separate ratings on cohort LEP share. The ratings are from 1 to 4, which translate to beginning, intermediate, advanced, and advanced high, respectively. Table 1.9 reports the coefficients of cohort LEP share under 4 different specifications as in other tables. Under specification 1, higher cohort LEP share is associated with lower English proficiency, as suggested by the summary statistics in Table 1.1. However, under the preferred specification, there appears to be no significant relationships between cohort LEP share and student English proficiency ratings, except for the 10% significance level on the student reading rating. Thus, I do not find evidence to support that high cohort LEP share improves students' English proficiency. Thus, differential English proficiency does not underlie the significant cohort LEP share effects that I find for math achievement, mainstreaming, and retention.

1.5.6. Instrumental Variable Estimates of the Effect of Exposure to LEP

Students on Student Achievement of Ever-LEP Students

In the above sections, I have looked at the reduced-form effect of cohort LEP share on student outcome. In this section, I estimate the direct effect of exposure to LEPs on the outcome of ever-LEP students using sample of 5th graders who have been in the district for at least a year. My cumulative measure of 5th graders' exposure to LEPs is the variable "average grade LEP share". The variable gives the share of classmates in a grade classified as LEP averaged over the grades the student has been through in the district. As discussed in an earlier section, this measure is endogenous in the sense that schools' unobservable factors affect both student outcomes and students being mainstreamed. Also, as argued above, cohort-to-cohort variation in LEP shares are idiosyncratic. Thus, if we can assume that cohort LEP share affects student outcomes only through "average grade LEP share", then we can estimate the effect of "average grade LEP share" on student outcomes using 2SLS.²⁷

Panels A and B in Table 1.10 provide the OLS and 2SLS results of exposure to LEPs on student achievement respectively. In Panel A, even though regressions include school fixed effects and school specific time trends, the estimates can still be biased due to the endogeneity of the measure of average grade LEP share. When I apply 2SLS using cohort LEP share as the instrument, I find that the effects of exposure to LEPs are positive and significant for math scores and are positive but insignificant for reading and language scores. The 2SLS estimate on the math scores is higher compared to the reduced-form estimate. Numerically, a 20-percentage point increase in exposure to LEPs increases math scores by 0.07 of a standard deviation. The 2SLS estimates are scaled by the first-stage estimates, which are provided in Panel C. Here, we see that cohort LEP share is positively correlated with "average grade LEP share". Specifically, a 20-

²⁷ As discussed in Section 1.3, it may be that cohort LEP share affects achievement through mechanisms other than exposure to LEPs. If the exclusion restriction is not valid, then the 2SLS estimates presented here are not consistent estimates of the effect of average LEP exposure. However, the reduced-form analysis presented above is still correct, and reveals the causal effect of increasing LEP share in one's cohort, even if we do not know the specific channels of this effect.

percentage point increase in cohort LEP share increases average grade LEP share by almost 8 percentage points.

1.6. Conclusion

Using data from an urban school district, I estimate the effect of cohort LEP shares on the academic outcomes of LEPs themselves. I rely on idiosyncratic variation on the shares of LEPs across cohorts within a school that is off the school-specific trend to identify the effect. I find that the cohort LEP share has a positive and significant effect on the student achievement in 5th grade, particularly on the math scores. Higher cohort LEP share also increases mainstreaming of LEPs while reducing grade retention.

While I do not have the data to look into all the mechanisms behind the estimated LEP peer effects, I have provided some evidence to support the conclusion that my estimates are not picking up effects from differential exposure to BE/ESL programs and peer achievement. I also find little to suggest that cohort LEP share affects students' English proficiency. Additionally, it appears that poorer schools or schools with higher cohort LEP shares mainstream students faster and retain students less in response to higher LEP shares, and that positive effect of cohort LEP share on math is higher in poorer schools. Moreover, I find that while the effect of cohort LEP share on math scores seems to be higher in larger school, its effect on mainstreaming and grade retention appears to be larger for smaller schools. The claim that economies of scale may be a channel through which cohort LEP share affects LEPs' outcomes can be questionable due to the differential effect depending on school enrollment. One of the channels that I am not able to examine is the teacher. Having more LEPs may allow teachers to focus their curriculum more to the needs of the LEPs. Not only that, in a district where the number of LEPs is consistently high, how a teacher teaches LEPs over time can matter as well. What current research suggests is that experience teaching LEPs is one of the most important factors in predicting a teacher's effectiveness with future LEPs (Master, Loeb, Whitney, and Wyckoff, 2011).

Overall, my findings suggest that interactions among LEPs and concentration of LEPs affect their educational outcomes. One can even argue that for the LEPs, learning with other LEPs can have such a positive effect on their achievement that despite the schools' policy of processing them through LEP services and elementary school faster, there is no evidence of negative effects on their achievement. The unexplored mechanisms of how cohort LEP share affects student outcomes may at the school level.

Table 1.1. Summary Statistics of the District's Ever-LEP 5th Graders from 1998-99 to 2009-10

	A. Full Sample		B. > Sample Median Cohort LEP Share		C. ≤ Sample Median Cohort LEP Share	
	Mean	SD	Mean	SD	Mean	SD
Dependent Variables - Stanford Achievement						
Test in 5th Grade						
Stanford Math Standardized Scores	0.016	(0.917)	-0.005	(0.896)	0.037	(0.936)
Stanford Reading Standardized Scores	-0.205	(0.878)	-0.252	(0.864)	-0.157	(0.889)
Stanford Language Standardized Scores	-0.153	(0.900)	-0.201	(0.881)	-0.105	(0.916)
Variables of Interest - Shares of LEP Students						
Share of LEP Students in a 1 st Grade - Cohort LEP Share	0.566	(0.197)	0.725	(0.087)	0.406	(0.137)
Average Shares of Currently LEP Students in a Grade	0.497	(0.178)	0.613	(0.116)	0.381	(0.151)
Control Variables						
Student Initial Achievement - Math	0.032	(0.995)	0.020	(0.988)	0.044	(1.002)
Student Initial Achievement - Reading	0.038	(0.983)	0.071	(0.973)	0.006	(0.992)
Student Initial Achievement - Language	-0.003	(1.003)	0.013	(1.010)	-0.018	(0.995)
Initial Math Exam was Stanford	0.210	(0.407)	0.151	(0.358)	0.268	(0.443)
Initial Reading Exam was Stanford	0.210	(0.407)	0.151	(0.358)	0.268	(0.443)
Initial Language Exam was Stanford	0.210	(0.407)	0.151	(0.358)	0.268	(0.443)
Female	.496	(0.500)	0.499	(0.500)	0.493	(0.500)
Economically Disadvantaged	0.949	(0.220)	0.964	(0.185)	0.933	(0.250)
Special Education	0.072	(0.259)	0.067	(0.250)	0.078	(0.268)
Gifted	0.161	(0.368)	0.166	(0.372)	0.156	(0.363)
Recent Immigrant	0.178	(0.382)	0.183	(0.386)	0.173	(0.378)
Native American	0.000	(0.016)	0.000	(0.015)	0.000	(0.016)
Asian	0.035	(0.183)	0.031	(0.172)	0.039	(0.193)
Black	0.010	(0.098)	0.008	(0.087)	0.012	(0.108)
Hispanic	0.946	(0.227)	0.956	(0.205)	0.935	(0.247)
White	0.010	(0.099)	0.005	(0.073)	0.014	(0.119)
Other Information						
Currently LEP	0.616	(0.486)	0.656	(0.475)	0.575	(0.494)
ESL	0.132	(0.338)	0.102	(0.302)	0.161	(0.368)
Bilingual	0.452	(0.498)	0.524	(0.499)	0.380	(0.485)
Number of Students in a School	536.772	(168.084)	565.174	(169.831)	508.409	(161.411)
Number of Currently-LEP Students in a School	279.804	(152.691)	348.593	(151.881)	211.108	(118.816)
Observations	44,436		22,203		22,233	

Notes: The full sample consists of all ever-LEP fifth graders in the district who have been in the district at least a year with no missing information. Panel B is restricted to observations with cohort LEP share above the full sample's median cohort LEP share, and Panel C is restricted to observations with cohort LEP share equal to or below that figure. Stanford/Aprenda scores are available from 1998-99 to 2009-10 and are standardized across grade and year over the entire student population in the district. Student initial achievement comes from student scores in either the Stanford (English) or Aprenda (Spanish) exam that the student took in the first year of entrance in the data. Variables "Initial (Math/Reading/Language) Exam was Stanford" are dummies indicating if the student takes the Stanford (rather than the Aprenda) in his/her first year in the data. A student is categorized as economically disadvantaged if the student is eligible for free or reduced lunch is classified, or is classified as "other disadvantage" (7.8 % of the observations). The summary statistics for the control variables reported in the table are for the sample with non-missing "Stanford Math standardized scores"; they are the similar to the samples with non-missing Reading and Language Scores.

Table 1.2. Variation in Cohort LEP Share

A. Raw Cohort LEP Share				
Observations	Mean	Standard Deviation	Minimum	Maximum
44,436	0.566	0.197	0.000	1.000
B. Residuals after Removing School Fixed Effects and Trends				
Observations	Mean	Standard Deviation	Minimum	Maximum
44,436	0.000	0.065	-0.432	0.423

Notes: The sample consists of all ever-LEP fifth graders in the district who have been in the district at least a year with no missing information. The residuals are calculated based on a regression of cohort LEP share on student initial achievement, student characteristics, shares of student characteristics in a grade and cohort, year dummies, school fixed effects and school-specific time trends.

Table 1.3. Reduced-Form Effects of Cohort LEP Share on Student Achievement

	Dependent Variable				
	Mean (S.D.)	(1)	(2)	(3)	(4)
A. Math	0.016 (0.917)	-0.099 (0.074)	-0.097 (0.137)	0.076 (0.063)	0.126** (0.056)
B. Reading	-0.205 (0.878)	-0.216*** (0.061)	-0.119 (0.107)	0.051 (0.062)	0.066 (0.060)
C. Language	-0.153 (0.900)	-0.191*** (0.072)	-0.121 (0.125)	-0.007 (0.067)	0.028 (0.061)
Observations	44,436	44,436	44,436	44,436	44,436

Notes: Sample consists of all ever-LEP fifth graders from school year 1998-99 to 2009-10 who have been in the district at least a year with no missing information. A student's cohort is a group of students who start 1st grade at the same school and year. For students who enter the district after 1st grade, which is about 15% of the sample, their cohort information comes from the cohort had they entered the district in 1st grade. Cohort LEP share is the share of students in a cohort who is LEP. Each coefficient reported comes from a separate regression. Specification used in columns (1) controls for year fixed effects, student initial achievement as well as student characteristics. Student initial achievement consists of student initial score in Math, Reading, and Language, whether the initially student take the Stanford exam, and the interactions of these indicators with the scores. Specification in columns (2) adds shares of student characteristics in a grade and cohort to the preceding column's specification. Specification in columns (3) adds campus fixed effects. Specification in columns (4) adds campus-specific time trends. Robust standard errors clustered at the campus level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.4. Reduced-Form Effects of Cohort LEP Share on Mainstreaming

A. Years Classified as LEP in Elementary School					
	Dep. Var. Mean (S.D.)	(1)	(2)	(3)	(4)
Grade 5: 1 st Grade	4.306	0.015	-0.207	-0.193*	-0.319***
Entrants Only	(1.460)	(0.100)	(0.161)	(0.105)	(0.099)
Observations	37,266	37,266	37,266	37,266	37,266
B. Probability of Being Mainstreamed					
	Dep. Var. Mean (S.D.)	(1)	(2)	(3)	(4)
Year 2	0.092	0.023	0.017	0.016	0.096*
	(0.289)	(0.023)	(0.035)	(0.048)	(0.054)
Observations	35,449	35,449	35,449	35,449	35,449
Year 3	0.137	0.015	0.021	-0.010	0.011
	(0.344)	(0.017)	(0.023)	(0.029)	(0.027)
Observations	35,449	35,449	35,449	35,449	35,449
Year 4	0.210	0.005	0.075	0.018	0.025
	(0.407)	(0.030)	(0.046)	(0.034)	(0.032)
Observations	35,449	35,449	35,449	35,449	35,449
Year 5	0.388	-0.091*	-0.056	-0.056	0.002
	(0.487)	(0.047)	(0.084)	(0.056)	(0.043)
Observations	35,449	35,449	35,449	35,449	35,449

Notes: In Panel A, sample consists of all ever-LEP fifth graders from school year 1998-99 to 2009-10 who have been in the data for at least a year and are 1st grade entrants. In Panel B, sample is restricted to the students who were LEP in first grade. I follow these students for the next 4 years. Mainstreamed is a dummy variable equal to 1 if the student is no longer identified as LEP. The dependent variable in Panel A is the number of years the student was classified as LEP. The dependent variable in Panel B is a dummy variable equals to zero. It becomes 1 when the student is declassified and thereafter. Each coefficient reported comes from a separate regression. Specification used in columns (1) controls for student initial achievement as well as student characteristics. Student initial achievement consists of student initial score in Math, Reading, and Language, whether the initially student take the Stanford exam, and the interactions of these indicators with the scores. Specification in columns (2) adds shares of student characteristics in a grade and cohort to the preceding column's specification. Specification in columns (3) adds campus fixed effects. Specification in columns (4) adds campus-specific time trends. Robust standard errors clustered at the campus level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.5. Reduced-Form Effects of Cohort LEP Share on Grade Retention

	Dep. Var. Mean (S.D.)	Probability of Being Ever-Retained			
		(1)	(2)	(3)	(4)
Grade 5: Full Sample	0.212 (0.409)	-0.018 (0.021)	-0.129*** (0.037)	-0.082** (0.034)	-0.096*** (0.035)
Observations	44,436	44,436	44,436	44,436	44,436
Grade 5: 1 st Grade Entrants Only	0.219 (0.414)	-0.023 (0.024)	-0.154*** (0.043)	-0.110*** (0.038)	-0.125*** (0.039)
Observations	37,266	37,266	37,266	37,266	37,266

Notes: Sample consists of all ever-LEP fifth graders from school year 1998-99 to 2009-10 who have been in the data for at least a year with no missing information. The second row is restricted to those who are 1st grade entrants. The dependent variable is a dummy equal to 1 if the student was ever retained from first to fifth grade. Each coefficient reported comes from a separate regression. Specification used in columns (1) controls for student initial achievement as well as student characteristics. Student initial achievement consists of student initial score in Math, Reading, and Language, whether the initially student take the Stanford exam, and the interactions of these indicators with the scores. Specification in columns (2) adds shares of student characteristics in a grade and cohort to the preceding column's specification. Specification in columns (3) adds campus fixed effects. Specification in columns (4) adds campus-specific time trends. Robust standard errors clustered at the campus level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.6. Heterogeneity in Effect of Cohort LEP Share by Student and School Characteristics

	Sex		Recent Immigrant Status		School's Average Economically Disadvantaged		School's Average Enrollment		Student Composite Initial Achievement	
	Female	Male	Recent Immigrant	Others	> Sample's Median	≤ Sample's Median	> Sample's Median	≤ Sample's Median	> Sample's Median	≤ Sample's Median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Math										
Observations	0.070 (0.078) 22,057	0.180** (0.074) 22,379	0.307** (0.133) 7,900	0.082 (0.062) 36,536	0.193*** (0.068) 21,895	0.056 (0.090) 22,541	0.150* (0.080) 22,050	0.117 (0.077) 22,386	0.175* (0.094) 22,218	0.047 (0.062) 22,218
B. Reading										
Observations	0.082 (0.076) 22,057	0.052 (0.081) 22,379	0.252* (0.138) 7,900	0.026 (0.060) 36,536	0.100 (0.081) 21,895	0.044 (0.088) 22,541	0.168** (0.082) 22,050	-0.022 (0.082) 22,386	0.054 (0.095) 22,218	0.078 (0.074) 22,218
C. Language										
Observations	0.072 (0.078) 22,057	-0.022 (0.084) 22,379	0.252* (0.138) 7,900	0.026 (0.060) 36,536	0.090 (0.083) 21,895	-0.043 (0.090) 22,541	-0.000 (0.090) 22,050	0.051 (0.082) 22,386	0.030 (0.099) 22,218	0.028 (0.071) 22,218
D. Years Classified As LEP										
Observations	-0.244* (0.131) 18,583	-0.411*** (0.125) 18,683	-0.511** (0.253) 4,293	-0.295*** (0.104) 32,973	-0.445*** (0.127) 18,297	-0.230 (0.146) 18,969	-0.218 (0.149) 18,127	-0.432*** (0.126) 19,139	-0.326** (0.129) 19,295	-0.375*** (0.132) 17,971
E. Ever-Retained										
Observations	-0.119** (0.047) 18,583	-0.131** (0.054) 18,683	-0.346*** (0.105) 4,293	-0.103** (0.042) 32,973	-0.135*** (0.049) 18,297	-0.119* (0.060) 18,969	-0.118* (0.064) 18,127	-0.138*** (0.051) 19,139	-0.051 (0.033) 19,295	-0.203*** (0.062) 17,971

Notes: Rows A, B, and C use data from the main sample, which consists of all ever-LEP fifth graders in the district who have been in the district at least a year with no missing information. Rows D and E is restricted to those in the main sample who are first-grade entrants. Each column uses a separate sample coming from the main or restricted sample defining by whether student is a female, a recent immigrant, whether the school's average economic disadvantage is above the sample's median, whether the school's enrollment is above the sample's median, and whether the student's composite initial achievement is above the sample's median initial achievement. A student composite initial achievement was calculated by taking an average of the student's standardized scores in Math, Reading and Language. Each coefficient reported comes from a separate regression that controls for student initial achievement, student characteristics, shares of student characteristics in a grade and cohort, campus fixed effects and campus-specific time trends. Robust standard errors clustered at the campus level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.7. Regressions of Student Initial ESL/Bilingual Status on Cohort LEP Share

	Initial Bilingual				Initial ESL			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	0.115***	0.019	-0.044	-0.046	-0.201***	-0.018	0.034	0.028
	(0.025)	(0.040)	(0.039)	(0.037)	(0.029)	(0.032)	(0.038)	(0.039)
Observations	44,436	44,436	44,436	44,436	44,436	44,436	44,436	44,436

Notes: The sample consists of all ever-lep fifth graders in the district who have been in the district at least a year with no missing information. Student initial Bilingual/ESL status is the Bilingual/ESL status of the student when he/she first enters the data. The dependent variable in Columns (1) to (4) is the dummy for initial placement into bilingual education, and the dependent variable in Columns (5) to (8) is the dummy for initial placement into ESL. The reported coefficient is for the variable "Cohort LEP Share", and each coefficient reported comes from a separate regression. Specification used in columns (1) and (5) controls for student initial achievement as well as student characteristics. Specification in columns (2) and (6) adds shares of student characteristics in a grade and cohort to the preceding column's specification. Specification in columns (3) and (7) adds campus fixed effects. Specification in columns (4) and (8) adds campus-specific time trends. Robust standard errors clustered at the campus level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.8. Examining the Effect of Cohort LEP Share on Student Achievement with Additional Controls

	Preferred Specification	Control for Peer Achievement
	(1)	(2)
Math	0.126** (0.056)	0.118** (0.055)
Reading	0.066 (0.060)	0.058 (0.060)
Language	0.028 (0.061)	0.019 (0.061)
Observations	44,436	44,283

Notes: Sample consists ever-LEP fifth graders who have been with the district at least a year with no missing information. Peer achievement is the average of peers' initial achievement. The reported coefficient is for the variable "Cohort LEP Share". Coefficients in the first column are from column (4) of Table 3, which are from the full specification with all controls, campus fixed effects and campus time trends. Initial peer achievement [average of peer achievement and share of peers who initially take Stanford (English) Exam] is added to regressions that produce coefficients in column 2. Robust standard errors clustered at campus level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.9. Relationship between Cohort LEP Share and Student English Proficiency

	(1)	(2)	(3)	(4)
A. Listening Rating	-0.187 (0.129)	-0.300 (0.213)	0.013 (0.099)	0.016 (0.093)
Observations	22,578	22,578	22,578	22,578
B. Speaking Rating	-0.174 (0.123)	-0.272 (0.203)	-0.026 (0.108)	-0.053 (0.101)
Observations	22,583	22,583	22,583	22,583
C. Writing Rating	-0.185* (0.102)	-0.242 (0.169)	-0.032 (0.116)	-0.078 (0.110)
Observations	22,514	22,514	22,514	22,514
D. Reading Rating	-0.074 (0.056)	-0.075 (0.079)	0.128 (0.079)	0.130* (0.077)
Observations	22,703	22,703	22,703	22,703
E. Composite Rating	-0.110* (0.061)	-0.155* (0.087)	0.073 (0.077)	0.065 (0.077)
Observations	22,299	22,299	22,299	22,299

Notes: The sample consists of all ever-lep fifth graders in the district who have been in the district at least a year with English proficiency ratings and other information available. The dependent variables are the ratings on students' language domains of listening, speaking, reading, writing and a composite rating that is determined from the ratings on the four language domains. The ratings are from 1 to 4, which translates to beginning, intermediate, advanced, and advanced high, respectively. The reported coefficient is for the variable "Cohort LEP Share", and each coefficient reported comes from a separate regression. Specification used in columns (1) controls for student initial achievement as well as student characteristics. Specification in columns (2) adds shares of student characteristics in a grade and cohort to the preceding column's specification. Specification in columns (3) adds campus fixed effects. Specification in columns (4) adds campus-specific time trends. Robust standard errors clustered at the campus level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.10. Effects of Exposure to LEP Students on Student Achievement of Ever-LEP Students

	A. OLS Estimates				B. 2SLS Estimates			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Math	-0.256*** (0.081)	-0.398*** (0.126)	0.047 (0.093)	0.161* (0.082)	-0.139 (0.100)	-0.146 (0.200)	0.187 (0.153)	0.326** (0.143)
B. Reading	-0.433*** (0.065)	-0.393*** (0.099)	0.058 (0.077)	0.195*** (0.072)	-0.303*** (0.078)	-0.179 (0.153)	0.125 (0.151)	0.171 (0.154)
C. Language	-0.449*** (0.077)	-0.496*** (0.121)	-0.086 (0.079)	0.105 (0.072)	-0.267*** (0.094)	-0.182 (0.180)	-0.018 (0.164)	0.073 (0.157)
Observations	44,436	44,436	44,436	44,436	44,436	44,436	44,436	44,436
					C. First-Stage (Coefficients for Cohort LEP Share)			
					0.715*** (0.029)	0.665*** (0.043)	0.407*** (0.012)	0.386*** (0.010)
Observations					44,436	44,436	44,436	44,436

Notes: The sample in all panels consists of all ever-LEP fifth graders in the district who have been in the district at least a year with no missing information. Each coefficient represents a separate regression. Panel A provides OLS estimates of the correlation between student average grade LEP share and student achievement. Panel B provides 2SLS estimates with 'average grade LEP share' as the endogenous regressor and 'cohort LEP share' as the identifying instrument. Variable 'average grade LEP share' gives the share of classmates in a grade classified as LEP averaged over the grades the student has been in the district. Cohort LEP share is the share of students in a cohort who is LEP. Panel C provides coefficients of cohort LEP share from regressions of student average grade LEP share on cohort LEP share. Specification used in column (1) & (5) controls for student initial achievement as well as student characteristics. Specification in column (2) & (6) adds shares of student characteristics in a grade and cohort. Specification in column (3) & (7) adds campus fixed effects. Specification in column (4) & (8) adds campus specific time trends. Robust standard errors clustered at campus level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 1.A1. Transition Matrix of Students Taking Aprenda Exam in 1st Grade

	Taking Aprenda in 1st grade	Fraction Taking Aprenda again in 2nd grade	Fraction Taking Aprenda again in 3rd grade	Fraction Taking Aprenda again in 4th grade	Fraction Taking Aprenda again in 5th grade
Math	75,597	0.754	0.580	0.286	0.007
Reading	75,692	0.755	0.580	0.286	0.007
Language	75,598	0.754	0.580	0.286	0.007

Notes: Sample consists of students from grade 1-5 from school year 1998-99 to 2009-10.

Table 1.A2. Shares of LEP Students Taking Stanford/Aprenda Achievement Exam

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Taking Stanford -Math	0.204	0.176	0.204	0.430	0.827
Taking Stanford - Reading	0.204	0.176	0.204	0.430	0.827
Taking Stanford - Lang	0.204	0.176	0.204	0.430	0.827
Taking Aprenda - Math	0.751	0.783	0.755	0.530	0.125
Taking Aprenda - Reading	0.752	0.783	0.756	0.530	0.126
Taking Aprenda - Lang	0.751	0.783	0.755	0.530	0.126

Notes: Sample consists of students from grade 1-5 from school year 1998-99 to 2009-10.

Table 1.A3. Transition Matrix of LEP Students to Mainstreamed Classes

	Number of students who were LEP in 1 st grade	Fraction being Mainstreamed
Grade 2	75,405	0.082
Grade 3	62,082	0.122
Grade 4	51,388	0.200
Grade 5	40,987	0.396

Notes: Sample consists of observations of students who are LEP in first grade, observing them in grade 2-5 from school year 1998-99 to 2009-10.

Chapter 2

Is Gifted Education a Bright Idea?

Assessing the Impact of Gifted and Talented

Programs on Students (with Scott A.

Imberman and Steven G. Craig)¹

2.1. Introduction

The focus of many school systems has been directed towards the low end of the achievement distribution as states have adopted accountability regimes and tried to meet the requirements of the federal No Child Left Behind Act (NCLB). As such, concerns have arisen that accountability pressures might force schools to shift resources away from high achieving students (Loveless, Farkas, and Duffett, 2008; Neal and Schanzenbach, 2010; Reback, 2008). Given these financial and regulatory constraints and the fact that over three million students in the US are classified as gifted and talented (GT), it is important to find out whether the services provided to these students are helpful. Hence, in this chapter we provide what are, to our knowledge, the first credibly causal estimates of the effects of GT programs on high achieving students.

¹ We would like to thank Aimee Chin, Dennis Epple, Jason Imbrogno, Chinhui Juhn, Brian Kovak, Jacob Vigdor and seminar participants at APPAM, AEEP, NBER Economics of Education Meeting, SOLE, Carnegie Mellon University, Georgetown University, University of Houston, University of Maryland, and the employees of an anonymous school district for all of their guidance and assistance in this project. Financial support from the University of Houston Small Grants program is gratefully appreciated. All errors remain our own.

GT programs might help high achieving students through peer effects, or additional resources.² While early research finds that ability grouping is correlated with higher achievement, many of these studies are likely biased due to unobserved characteristics of students, such as motivation, that simultaneously lead students to be successful and to be grouped in high ability classrooms.³ Recently, research has tried to address the bias issue in ability grouping, but with mixed results (Argys, Rees and Brewer, 1996; Betts and Shkolnik, 2000; Epple, Newton and Romano, 2002; Figlio and Page, 2002; Duflo, Dupas and Kremer, 2011). Our work addresses the reduced-form impact of GT programs, but uses two unique strategies for overcoming bias - a regression discontinuity design embedded within the eligibility requirements, and an analysis of lotteries for entrance into two premier GT magnet programs. This combination of approaches, although asking different but nonetheless closely related questions, spans the distribution of GT students, as the RD examines students marginally eligible, while the lottery examines relatively strong students even within the GT population.

Specifically, we utilize a universal GT evaluation in a large urban school district in the Southwest United States (LUSD) where, since 2007, all fifth grade students have been evaluated to determine eligibility for GT services starting in sixth grade. Eligibility is determined by two well-defined cutoffs on an index score that is based on achievement tests, a non-verbal ability test, grades, teacher recommendations, and socio-economic status. We exploit these cutoffs to set up a regression discontinuity (RD) design whereby students who score just above the cutoffs are compared to those who score just below. Under certain conditions, for which we provide evidence that this analysis meets, our estimates provide the causal impact of enrolling in a GT

² Additional resources can include special training of teachers. The curriculum also changes; conversations with district officials in the district we study suggest that the GT curriculum includes more detail and goes more in-depth into topics, rather than cover the regular curriculum more quickly or add additional topics (increase breadth).

³ See Kulik and Kulik (1997) for a review.

program on achievement for students on the margin of eligibility relative to enrolling in a “regular” program.

The second research strategy we employ uses randomized lotteries that determine admission to two middle schools with over-subscribed magnet GT programs. Conditional on meeting the district-wide GT eligibility requirements and completing an application, students not in the attendance zones are randomly offered admission to the district’s premier magnet schools. This allows us to examine achievement and attendance differences between students who win the lottery and attend the magnet GT schools, and those who lose the lottery and attend other “neighborhood” GT programs.⁴ This analysis provides evidence on the impact of the extra inputs, including peers, because we provide evidence that the premier GT schools generate a more intensive treatment along observable measures than the neighborhood GT programs.

The combination of our two research strategies provides an unusually broad look at how GT programs affect student achievement. While our RD approach uses students that are marginally eligible, since there are several measurement dimensions for GT eligibility our work covers a broader array of students than would be typical in many RD applications. The lottery application captures relatively high achieving students even among those GT eligible, albeit with a relatively small sample. Further, while each approach has different comparison groups, in both cases we decisively show there are additional inputs into the education production function.⁵ Specifically, the marginally eligible students in the RD analysis receive a different curriculum, and stronger peers. Students that win the lottery, while receiving the same basic curriculum as the lottery losers, nonetheless get stronger teachers and peers. The combination of the two

⁴ Students that lose the lottery can attend GT programs in their local school, other magnet schools (based on other specializations), or charter schools. The GT programs in these other schools are called neighborhood programs, because they are not designed to attract GT students from other attendance zones.

⁵ While students who lose the lottery also have the opportunity to attend a non-lottery magnet, the vast majority of those who stay in the district attend a neighborhood school program.

approaches, therefore, has the strength of measuring the effectiveness of additional school inputs at different points in the student quality distribution.

Several recent papers examine the impact of elite schools, but with mixed results (; Hoekstra, 2009; Clark, 2010; Jackson, 2010; Abdulkadiroglu, Angrist and Pathak, 2011; Fryer and Dobie, 2011). Nonetheless, to our knowledge, only Bhatt (2010) and Murphy (2009) specifically study the effect of GT programs on achievement, although Davis, Engberg, Epple, Sieg and Zimmer (2010) find that higher income parents are more likely to stay in public schools when their children are eligible for GT programs. Bhatt finds significant improvements in math achievement, although her instrumental variables methodology suffers from weak instruments. Murphy (2009) finds little math or reading improvement from being identified as GT, although these results may suffer from bias if trends in achievement determine program entry. Our work offers a wider scope of inquiry, and further offers the two distinct identification strategies. Thus, in our view, our study is the first to establish credibly causal estimates of the impacts of GT programs on student achievement.

Our RD analysis shows that students exposed to GT curriculum for the entirety of 6th grade plus half of 7th grade exhibit no significant improvement in achievement as measured by standardized test scores. This is despite substantial increases in average peer achievement on the order of one-fourth to one-third of a standard deviation, as well as being provided an advanced GT curriculum. We also find no significant improvements within student subgroups. While our estimates are local to the RD margin we note that, due to the multiple factors that go into the eligibility index, students near the discontinuity show wide variation in achievement levels, leading to a more generalizable interpretation than most RD designs. For example, students who

score precisely at the eligibility threshold range from 45 to 97 national percentile rankings in reading and between 55 and 98 percentiles in math on the 5th grade exams.⁶

The lottery results that compare students in the two premier GT magnet schools to other GT students also show little improvement in 7th grade achievement with the exception of science scores. This is despite improvements in mean peer achievement on the order of one standard deviation, and higher quality teachers as measured by teacher fixed effects. The advantage of the lottery analysis is that, despite that the treatment differs from that studied in the RD (the RD analysis measures the impact of GT enrollment itself while the lotteries identify differences in the intensity of treatment) the students from the lottery are on different segment of the achievement distribution.⁷ Nonetheless, despite the substantial difference in peer and teacher quality, we find lottery winners only out-perform losers in science.

We believe these results are somewhat surprising, although there is evidence in the literature that corroborates our findings. That marginal GT students are unaffected by the program might indicate that the school district sets the entry cutoff at an appropriate level so that there is no marginal benefit of enrolling. Two factors, however, work against this explanation. First, if this were the case we would expect to see strong positive effects from the GT magnet lotteries. Second, these results are consistent with those found by Abdulkadiroglu, Angrist and Pathak (2011) for elite magnet high schools in Boston and New York and Fryer and Dobbie (2011) in New York. Another mechanism we explore is the role of peer effects. If peer effects follow a monotonic model whereby being surrounded by higher achieving students improves one's own achievement, as found in Imberman, Kugler and Sacerdote (forthcoming), better peers should lead to achievement gains. For a marginal GT student, however, the peer effect may not

⁶ The inter-quartile ranges are 78 to 92 for reading and 83 to 94 for math. Additional measures for GT qualification include an intelligence test, grades, teacher recommendations and socioeconomic disadvantages.

⁷ While the average GT student in 7th grade scores 1 standard deviation above the mean 7th grade student in the district, the lottery participants score 1.6 standard deviations above the mean.

necessarily be positive. That is, a marginal GT student is likely to go from being near the top of the regular class to being near the bottom of the GT class. Thus it is possible that an invidious comparison (IC) model applies, such as proposed by Hoxby and Weingarth (2006), whereby students are demoralized by reductions in their relative ranking.

Further complicating the impact of peer effects in this context is that how the teacher targets her instruction (e.g. to the median, bottom or top student) can affect the benefits to the marginal student (Duflo, Dupas and Kremer, forthcoming). Further, Cicala and Fryer (2011) argue that the impact of moving students into an environment with higher achieving peers depends on where the student is in the achievement distribution. We therefore examine evidence using both course grades and the within class ranking of students, and while we find that both fall when a student enrolls in GT, we do not find evidence of an interaction effect between treatment and prior achievement that would suggest IC plays a large role.

Finally, our results could also be explained by the nature of achievement testing. It is possible that standardized tests do not adequately reflect improvements in creativity and critical thinking, attributes that GT programs purportedly address, and hence this may not be the best outcome in this context. Unfortunately, while less than ideal, the use of standardized testing is necessary as the students in our study have not progressed enough in school to analyze impacts on longer term outcomes that may reflect this, such as college going or college entrance exam scores.

2.2. The Gifted and Talented Program in LUSD

LUSD is a large school district in the Southwestern US with over 200,000 students. The district is heavily minority and very low income, where the minority population is mostly Hispanic. The district claims that GT classes differ from others by going deeper into the material, as opposed to increasing the breadth of topics. Much of this is done through the use of special

projects that reflect the curriculum and cut across subject matters. While teachers and principals ultimately have substantial control over what is taught, the state does provide some recommended lessons. One example is a project that has students conduct independent research on an issue in health. Another has students conduct statistical analyses of everyday life events. Thus the GT program is geared towards enhancing creative and critical thinking compared to its regular program.

Panel A of Table 2.1 provides characteristics of the GT and non-GT students we study. It shows that gifted students are less likely to be economically disadvantaged, more likely to be white, less likely to have limited English proficiency, and they perform better on cognitive and non-cognitive outcomes than non GT students.⁸ To be identified as GT in LUSD, a student must meet the eligibility criteria set forth in the “gifted and talented identification matrix.” The matrix for GT in 2008-09 is provided in Figure 2.1. It converts scores on standardized tests, Stanford Achievement Test for English speaking students and the Aprenda exam for Spanish speaking students with limited English proficiency, plus scores on the Naglieri Non-verbal Abilities Test (NNAT), average course grades, teacher recommendations, and indicators for socio-economic status into an index score we call “total matrix points.”⁹

Students can meet eligibility requirements in one of two ways. The first is having 56 total matrix points, including at least 16 points from the Stanford Achievement Test or Aprenda and 10 points from the NNAT.¹⁰ Alternatively, students can qualify by having 62 total matrix points regardless of Stanford, Aprenda and NNAT scores. During 5th grade all students are evaluated for

⁸ Schools in LUSD also have a monetary incentive for attracting gifted students as LUSD provides a funding boost of 12% over the average daily allotment for a regular student.

⁹ For socioeconomic status, students get 5 extra points (out of a maximum of 108) for having limited English proficiency, being classified as special education or being classified as economically disadvantaged. Students who are members of a minority group get a further 3 point bonus.

¹⁰ Students can reach 16 points from the Stanford Tests through scores in four subjects. For example a student would qualify by being in the 90th percentile in math and the 80th percentile in reading regardless of science and social studies scores. Alternatively a student could score in the 80th percentile in all four exams. See Figure 2.1 for details on the conversion of test scores to points.

GT, including those who participated in the GT program in elementary school.¹¹ This selection framework allows us to model qualification across the eligibility boundary by using a fuzzy RD methodology without concern for selection on evaluation. The fuzzy RD design is because even among students who qualify, not all end up being classified as GT as parents are allowed to opt-out of the program, or students may enroll and then withdraw.¹²

Table 2.1 also shows means from the lottery sample in panel B. The students in the lottery are significantly stronger academically than the average GT student in panel A. These differences distinguish the lottery sample from the marginally eligible students in the RD. For example, lottery participants in Panel B are shown to score about 1 standard deviation higher on standardized tests in 5th grade than the mean overall GT student from Panel A.¹³ Lottery participants are also more likely to be white, and not on subsidized school lunch. Attrition bias is a potential factor, since 18.8% of the 542 students that enter the lottery are not in the school district by 7th grade. In fact, most of these students leave the sample before 6th grade, and further the leavers are different than the lottery winners as shown in the last two columns. We address potential attrition bias in two ways – by reweighting the sample to look like the pre-lottery sample on observables, and through the use of a bounding analysis proposed by Engberg, Epplé, Imbrogno, Sieg and Zimmer (2010).¹⁴

2.3. Model and Specification

¹¹ Elementary students must re-qualify in 5th grade to maintain their classification in middle school. Students who qualify for GT in middle or high school generally keep their status through graduation, although they can be removed from GT if they perform poorly. Students can qualify in later years at the school's discretion, they are omitted from our data here.

¹² Another reason a student may not show up in the data as GT is if his or her school does not have enough GT certified teachers to provide the required services. This is very rare, however, as only 2 of the 41 traditional middle schools in LUSD had no GT students in 7th grade in 2009-10. Further, it is possible to qualify for GT in 7th grade, but this possibility does not affect our estimates as the RD is based on qualification after 5th grade assessment, and the control is based on not being in GT in 6th and 7th grade.

¹³ Throughout this chapter we standardize scale scores from each exam within grade and year across the district.

¹⁴ We also test for selective attrition in the RD sample but find little evidence to suggest it is a problem.

2.3.1. GT Program Evaluation Using Regression Discontinuity

The objective of the RD analysis is to estimate a local average treatment effect of providing gifted services to students who are on the margin of GT qualification. Figure 2.2 shows GT identification two years after evaluation (7th grade) as a function of the students' matrix points. The gradual increase up to 28% at the first cutoff (of students with a matrix score of 56) reflects missing matrix components, qualifying in 7th grade and the district's appeal process. Upon reaching the first threshold, GT enrollment jumps to 45 percent. Enrollment increases further at a steep rate between the two cutoffs, hitting 79% at the second cutoff (62 matrix points). After reaching the second cutoff, GT enrollment slightly increases further to 82 percent.

Given that the increase in GT over this range is steep but not discontinuous, we convert the two thresholds into a single cutoff. To do this we map components of the matrix scores into three-dimensional space as shown in Figure 2.3.¹⁵ Each axis reflects one of the three portions of the matrix score that determines eligibility – NNAT points, Stanford/Aprenda points, and other points, which includes socio-economic status, grades, and teacher recommendations. Students who are on or above the surface are eligible for GT while those below or behind it are ineligible. We then take the Euclidean distance from each student's total matrix points to the closest integer combination on the surface.¹⁶ The resulting value, the distance to the qualification threshold, equals zero if the student just barely qualifies for GT. Figure 2.4 shows GT enrollment as a function of Euclidean distance from the threshold. Students just below the cutoff have a 25% likelihood of being in GT while students just above have a likelihood of about 79 percent.¹⁷

¹⁵We also estimate models using each of the two cutoffs individually for subsamples affected by each cutoff and find similar results.

¹⁶The Euclidean distance is measured as $Distance_i = \sqrt{(Stanford_i - Stanford_s)^2 + (NNAT_i - NNAT_s)^2 + (Other_i - Other_s)^2}$ where i refers to the student's own score and s refers to the closest integer combination on the surface. We thank Jake Vigdor for first suggesting this method to us. See Fig 2.1.

¹⁷Note that by construction the distance measure has an empty mass between 0 and 1, and between -1 and 0, since the smallest distance to another integer point is 1.

The fuzzy RD model is based on a two-stage least squares regression within a range of values that includes the cutoff (Hahn, Todd and Van der Klaauw, 2001; Lee and Lemieux, 2010). We generally use ten distance units below and above the cutoff for our bandwidth since the relationships between distance and the achievement outcomes are close to linear over this range, allowing us to use a linear smoother. We show later that our results are not sensitive to the choice of bandwidth. Hence, we estimate the following two-stage least squares (2SLS) model:

$$(1) GT_{i,t+k} = \delta + \gamma Above_{it} + \rho_1 Dist_{it} + \rho_2 Dist_{it} \times Above_{it} + \Omega X_{it} + \mu_{ijt+k}$$

$$(2) Y_{i,t+k} = \alpha + \beta GT_{i,t+k} + \lambda_1 Dist_{it} + \lambda_2 Dist_{it} \times Above_{it} + \Phi X_{it} + \varepsilon_{ijt+k}$$

where $Above_{it}$ is an indicator for whether student i in year t has a distance measure at or above the cutoff, $Dist$ is the Euclidean distance of the student's matrix score to the eligibility cutoff, and X is a set of pre-existing (5th grade) observable characteristics which includes the 5th grade dependent variable (e.g. lagged achievement), gender, ethnicity, gifted status, economic disadvantaged status, and LEP status. GT is an indicator for whether the student is enrolled in a GT program in year $t + k$ and Y is test scores, attendance, or disciplinary infractions in that year. Since students are tested in January of each year, we focus on outcomes in the second year after evaluation (7th grade) as assessment in the first year will only provide five months of program exposure, although we provide estimates for 6th grade outcomes in the online appendix.

2.3.2. GT Magnet Evaluation Using School Lotteries

LUSD has two middle schools with GT magnet programs that are over-subscribed, and as a result the district uses lotteries to allocate spaces.¹⁸ While the losers of the lottery have the opportunity to receive GT services in other schools, the magnet schools are considered to be

¹⁸ There are 8 middle schools with GT magnet programs in total (out of 41 traditional middle schools), but only two are over-subscribed. By seventh grade, of the 109 lottery losers that stay in LUSD, 21 enroll in one of the lottery magnet schools, only 5 attend one of the other six GT magnet programs, and the remainder attend a neighborhood GT program. Conversely, of the 265 lottery winners, only 3 attend one of the other six GT magnets in 7th grade.

premium schools due to their large GT populations and advanced curricular focus.¹⁹ The lottery thus allows us to examine impacts on a different segment of the GT student quality distribution.

Our analysis compares the performance of students who win the lottery and attend one of the magnet GT programs to those who lose the lottery and either attend a neighborhood GT program in the district, a magnet school based on a different specialty, or a charter school. Hence in the lottery sample we estimate the following 2SLS model conditional on applying for admission to a magnet program with a lottery:

$$(3) GTMagnet_{ijt+k} = \delta + \gamma Admitted_{ijt} + \mathbf{\Omega} \mathbf{X}_{it} + v_j + \mu_{ijt}$$

$$(4) Y_{ijt+k} = \alpha + \beta GTMagnet_{ijt+k} + \mathbf{\Phi} \mathbf{X}_{it} + \eta_j + \epsilon_{ijt}$$

where *GTMagnet* is an indicator for attending any GT magnet program, including those that do not hold a lottery, *Admitted* is an indicator for being offered a slot at a program with a lottery, and *X* is a set of student level controls. Finally, since each school holds separate lotteries we include v_j and η_j in the model as lottery fixed-effects.²⁰

One caveat to the lottery is that students with an older sibling in the school are exempted from the lottery and automatically admitted. Unfortunately, LUSD was unable to provide data on siblings, but we believe siblings have a negligible impact on our estimates. First, siblings still need to apply and qualify for GT based on their matrix points. Second, our lottery sample is very well balanced between “winners,” including those accepted under the sibling rule, and “losers,” thus indicating selection effects are unlikely. Finally, even if older siblings potentially offer advantages, lottery losers may have older siblings in the school they ultimately attend.

¹⁹ One of the two lottery schools also has an attendance zone. GT students from the attendance zone bypass the lottery, hence we drop from our sample any student zoned to that school.

²⁰ Since we focus only on one cohort, 5th graders in 2007-08 (who are in 7th grade in 2009-10), there is a single lottery fixed-effect indicator in each regression. Models with 6th grade outcomes, and hence two cohorts of students, have three indicators. Also note that coding *GTMagnet* for students who attend a non-lottery GT magnet as zero instead of one has no effect on the results.

2.4. Regression Discontinuity Estimates of GT Impacts

2.4.1. Data

Our data consists of the administrative records from 2007-08 to 2009-10. While we have data for universal assessments conducted in 2006-07, many schools were given exemptions from the new rules that year in order to allow for an orderly transition to the new system. As such, we start our sample in 2007-08, the second year of the mandatory GT assessment, and examine outcomes through the 2009-10 school year. For outcomes we use scale scores standardized across LUSD within grade and year on the Stanford Achievement Test, as well as attendance rates and counts of disciplinary infractions warranting an in-school suspension or more severe punishment. The test results are in standard deviation units relative to the district-wide distribution in a grade for math, reading, language, science and social studies exams.²¹ After restricting the sample to a 20 unit band around the cutoff, we look at achievement of approximately 2,600 students in one 7th grade cohort and 5,500 students in two 6th grade cohorts who were evaluated for GT in 5th grade.²²

2.4.2. Tests of Validity of RD Design

In Figure 2.5, we provide density plots around the eligibility discontinuity showing differences in density around the discontinuity are similar in size to changes at other parts of the distribution, suggesting that bias in our results from manipulation is unlikely to be occurring (Lee and Lemieux, 2010).²³ In Table 2.2 we provide tests of discontinuities in pre-existing (5th grade)

²¹ While some LEP students are evaluated using the Aprenda exam, a Spanish-language alternative to the Stanford Achievement Tests, only 0.5% of 5th grade students in LUSD take only the Aprenda exam and hence have no Stanford scores. Thus, we drop students who only have Aprenda scores.

²² Within this bandwidth the total matrix scores have an inter-quartile range of 48 to 65 with a minimum of 39 and a maximum of 79.

²³ Ideally one would like to conduct McCrary's (2008) test. Since there are no observations between 1 and 0 or -1 and 0 (see note 16) and there is positive mass between integers further out, this could mistakenly generate a positive result. Hence, instead we test for discontinuities at the two cutoffs in the total matrix points distribution to check for manipulation. In both cases the test is statistically insignificant.

student characteristics.²⁴ We find no discontinuities in columns 1-14 for race, gender, LEP status, prior gifted status, special education status, eligibility for free or reduced-price lunch, disciplinary infractions, attendance rates, and achievement with the exception of math.²⁵ Given that math is the only covariate that is significant we believe this to be a spurious result. Nonetheless, since achievement is highly correlated over time we correct for this by providing results both with and without controls that include the lagged (5th grade) dependent variable.

In column (15) we test whether there is any difference in whether a component of the matrix is missing, and find no such evidence. The next two columns address whether teachers manipulate evaluations for students at the qualification threshold.²⁶ If this were the case we would expect to find a discontinuity in the teacher scores, or in the matrix points the student gets from the teacher. We find no statistically significant discontinuity in either measure of teacher recommendation. Nonetheless, below we provide an additional specification test to check for potential bias from teacher manipulation through their recommendations.

Finally, in columns (18) through (20) we test whether there is a discontinuous likelihood of being enrolled two years after evaluation. Given that Davis, et al. (2010) find evidence that high income students are more likely to stay in public schools if identified as GT, we check if such a phenomenon occurs in LUSD. Additionally, a discontinuity in attrition would be a marker

We also provide graphical evidence on the distribution of matrix points in Online Appendix Figure 1.

²⁴ A related concern is that laid out by Barreca, Guldi, Lindo and Waddell (2010), that heaping in running variables could lead to biased estimates if the heaps are correlated with unobservables and the bandwidths are small enough so heaps are concentrated only on one side of the cutoff. We do not find evidence of heaping in matrix scores. By construction some heaping will occur, however, in the transformation from matrix scores to Euclidean distances. Our bandwidths, however, are wide enough to include substantial observations both at heaping and non-heaping points on both sides of the cutoff. We will also show later that our results are quite robust to choice of bandwidth.

²⁵ Online Appendix Figures 2 – 4 provide graphical representations of these results. Tests that do not condition on appearing in the data in 7th grade are similar and are provided in Online Appendix Table 1, along with tests using the 6th grade sample. These are also similar for all measures except for females which show a small but statistically significant increase. In Online Appendix Table 2 we provide estimates without clustering of standard errors. These show no change in the significance levels of the estimates.

²⁶ Although teacher recommendations are due before achievement scores are calculated, district officials informed us that many teachers submit their recommendations late.

for potential attrition bias in the estimates. Nonetheless, we find no statistically significant change in the likelihood of enrollment at the discontinuity regardless of economic status. Hence, given the totality of results described here we see little evidence that GT qualifications were manipulated in a way that would violate the assumptions underlying the RD methodology.

2.4.3. Results

We present both graphical and econometric evidence that GT exposure has no discernible effects on achievement test scores after a year and a half of exposure. We also present extensive sensitivity analysis that shows the results are robust to a variety of alternative specifications.

Figure 2.6 presents the initial reduced-form results for all five achievement tests. These achievement test results are from 7th grade, and thus reflect eighteen months of GT program exposure. We use a bandwidth in our regressions that includes students within ten distance units of the GT qualification boundary.²⁷ The coefficient estimates presented in the first panel of Table 2.3 as well as the results provided in Figure 2.6 show that there is no improvement in reading, language, science or social studies Stanford scores from GT participation, and that there is a negative and significant point estimate for math without individual controls.²⁸

Panel B of Table 2.3 provides estimates from our preferred specification of equation (2) that contains student level controls measured during 5th grade, including the lagged dependent variable, race, gender, economic disadvantage, LEP status, and gifted status. In this panel, all of 2SLS estimates are close to zero while math, reading and social science are negative and all t-statistics are below one. Drawing 95% confidence intervals around the estimates, we can rule out modest positive impacts of GT on marginal students of more than 0.06 standard deviations (sd) in

²⁷ Online Appendix Figure 5 shows that our Euclidean distance measure correlates very well with total matrix points.

²⁸ Note that 89% of students in GT in 7th grade are also in GT in 6th grade and, conditional on remaining in LUSD, 4% of students who are in GT in 6th grade are not GT in 7th grade.

math, 0.09 sd in reading, 0.15 sd in language, 0.13 sd in social studies and 0.23 sd in science. The point estimates themselves, however, clearly suggest a zero effect.²⁹

In columns (6) and (7) we examine impacts on non-cognitive outcomes, disciplinary infractions and attendance rates. While there is no effect on disciplinary infractions, we do find a marginally significant negative effect on attendance rates. The drop in attendance rates of 1.1 percentage points is equivalent to attending school two fewer days in a 180 day school year. As we demonstrate below, however, this estimate is sensitive to the specification.

Panel C presents results that correct for the possibility of teacher manipulation. Even though our earlier tests did not suggest a problem, we are especially concerned because an administrator acknowledged that time deadlines are lax for recommendations, leaving opportunities for teachers to “top up” the scores of marginal GT students. We address this possibility by replacing a student’s matrix points with a synthetic matrix score if the points from a teacher recommendation are potentially pivotal. That is, for students whose other matrix components place them within 10 points of the cutoff, the teacher recommendation, with a maximum of 10 points, is potentially determinative.³⁰ Thus for these students we replace their total matrix score with the predicted value from a regression using the full 5th grade sample of total matrix points on all matrix components excluding teacher points:

$$(5) TotalPoints_i = \alpha + \beta StanfordPoints_i + \gamma NNATPoints_i + \delta ObstaclePoints_i + \eta GradePoints_i + \epsilon_i.$$

²⁹ Online Appendix Table 3 provides results for 6th grade. These are similar to those for 7th grade and are more precise due to the addition of an extra year of data. For social studies we can rule out effects of 0.15 sd while for other tests we can rule out impacts of 0.09 sd and higher. Appendix Table 4 provides results with the lagged dependent variable but without the other covariates. These are similar to the results in Panel B of Table 2.3. Finally Appendix Table 5 shows the results to be robust to the inclusion of middle school fixed-effects.

³⁰ While teacher recommendations would not be pivotal for students within 10 points above the cutoff, only adjusting scores on one side of the cutoff could introduce bias, particularly if teachers only have information on some of the components at the time they make their recommendations. Hence, we replace scores within 10 points above the cutoff with the synthetic scores as well.

where *TotalPoints* is the student’s final score on the GT qualification matrix, *StanfordPoints* are the number of matrix points received on Stanford Achievement Tests, *NNATPoints* are matrix points from the non-verbal abilities test, *ObstaclePoints* are matrix points from socioeconomic status, and *GradePoints* are matrix points from the student’s average grades in 5th grade. We convert these “synthetic matrix points” to Euclidean distances from the eligibility surface, thus purging the teacher component and any potential manipulation from the matrix scores.

The results using the synthetic scores including all controls are provided in panel C. Since we are essentially adding measurement error to the first stage, the cutoff instruments are considerably weaker. The estimates nonetheless show a clear discontinuity that is highly significant.³¹ The 2SLS results show negative and insignificant effects for math, reading, language and social science while science estimates, although positive, are very close to zero. Discipline results are also similar to those in panel B. On the other hand, while attendance estimates are statistically insignificant, they turn positive. Given the potential influence of attendance on teacher recommendations, further RD analyses on attendance should be interpreted with caution. Nonetheless, since the results for all other outcomes are consistent with the estimates in panel B, the baseline model with controls is our preferred specification.³²

To test for heterogeneity in program impacts across student characteristics, Online Appendix Table 6 provides 2SLS estimates for 7th grade for various student sub-populations. In general, we find little evidence of differences by gender, race/ethnicity, economic status or prior gifted status. The only distinction is that the attendance estimates are more negative for women and black students and positive for white students, but we are cautious in drawing interpretations from this given the results in panel C of Table 2.3.

³¹ In Online Appendix Figure 6 we provide a graph of the first stage for the synthetic teacher scores.

³² We also estimated models that drop all students where their entry into GT could be impacted by teacher scores. While the estimates were very imprecise results were qualitatively similar to those in panel C of Table 2.3.

Table 2.4 reports sensitivity analyses using the preferred model from Table 2.3, where we find that our estimates hold regardless of whether we add middle school fixed-effects, limit the data to observations with no missing matrix components, use smaller or wider bandwidths, or conduct local linear regressions with optimal bandwidths determined by leave-one-out cross validation. We also estimate models where, instead of using the distance index, we restrict the samples to students who score high on Stanford and NNAT and hence are eligible for the 56 point cutoff (row 5) and students who score lower on these tests and hence are eligible for the 62 point cutoff. We use the raw matrix score, instead of the Euclidean distance, as the forcing variable in these regressions. In both subsamples, the results are similar to our baseline model.

When we use a quadratic smoother as the functional form, however, the estimates show significant improvements in language and science achievement scores. These results become insignificant using a cubic smoother due to larger standard errors. Further inspection, however, suggests that this estimate is being driven by excessive curvature at the discontinuity. Online Appendix Figures 7 and 8 show that the quadratic estimates are mainly driven by random variation at the discontinuity and hence, tend to overestimate what appears to be the true impact. As such, we believe a linear smoother captures the correct estimates.

A further possibility is that the lack of positive effects may be due to ceiling effects where many students have scores close to or at the maximum. To address this, in Online Appendix Figures 9 – 13 we provide distribution plots of raw scores on each of the 7th grade Stanford Achievement Tests for students with Euclidean distances between -10 and 10. For these marginal GT students, in all cases we find the mass of the distribution is centered far from the maximum score. For example, in math the modal score is 62 out of 80 while it is 67 out of 84 for reading, leaving substantial room for improvement.

Another potential reason for not finding an effect of GT services on student outcomes is that there may be little treatment on students. In Table 2.5, however, we illustrate the extent to

which entering GT generates a measureable treatment. We estimate the impact of GT on peer achievement, where a student's peers are determined by other students in a grade-teacher-course cell,³³ school choices, teacher quality and enrollment in "Vanguard" classes which are pre-Advanced Placement classes with advanced curricula targeted to gifted students.

Teacher quality is measured through the use of teacher fixed-effects using a procedure in the spirit of Kane and Staiger (2008). Specifically, we use teacher-student linked achievement data for middle school students from 2007-08 to 2009-10 to estimate the following model:

$$(6) Y_{ijkt} = \alpha + \delta Y_{ijkt-1} + \Omega X_{ijkt} + \Phi Z_{ijkt} + \gamma_k + \delta_j + \epsilon_{ijkt}$$

where Y is achievement in a given subject for student i , and X is a set of student level controls for economic status, gender, race/ethnicity, special education status, LEP status, and grade by year fixed effects. Z is a set of controls for mean peer achievement (defined at the grade-course-teacher level), while δ_j is a set of school fixed effects. Teacher quality is measured using the teacher fixed effect terms γ_k .³⁴ Kane and Staiger (2008) show that a similar framework closely replicates the results from a randomized allocation of students to teachers.³⁵

³³ Ideally one would like to use the actual classroom as the peer group. Unfortunately specific course section data are not available. To test the extent to which this is an issue, in Online Appendix Table 7 we sort students into synthetic classrooms of at most 35 students under the assumption that students are tracked by their 5th grade achievement in the given subject (row (i)) or randomly (row (ii)). With the exception of math in 7th grade the estimated change in peer achievement is similar to those found in Table 2.5 under both assumptions.

³⁴ We estimate this model such that each observation is assigned a weight equal to the teacher's share of classes taught to a student in a given subject. For example if a student takes a class in US history and another class in geography, then the student will have two observations in the social studies regression, one for each class, with a weight of ½ for each observation. Additionally, since the Stanford exams are given in January, we assign to each student the teachers they had in the spring of the previous academic year and the fall of the current academic year.

³⁵ Our model diverges from Kane and Staiger (2008) in two key ways. First, they use a random effects rather than fixed effects framework. We prefer the latter as it allows for weaker identification assumptions. Second, they utilize a Bayesian smoother that adjusts estimates for teachers with few observations towards the mean. While this strategy is important when trying to identify the influence of teachers on students, it is inappropriate in our context testing whether GT students receive higher quality teachers, as teachers with fewer observations will tend to be younger and less experienced, hence pushing their estimates to the mean would give us biased measures of actual teacher quality.

Table 2.5 shows in columns (1) to (5) that peer achievement is between 0.24 and 0.35 standard deviations higher for GT students relative to non-GT students.³⁶ The table also shows that GT students are more likely to enroll in Vanguard classes and attend a GT magnet program. Interestingly, most school switching does not come from students leaving their zoned school for GT magnets. Rather, students switch from schools other than their zoned school – mostly non-GT magnets – to the GT magnets. Finally, and perhaps surprisingly, GT students do not appear to get assigned better teachers as measured by teacher fixed effects. This may be because in many schools both GT and non-GT students can access advanced classes taught by the same teacher. Nonetheless, the change in peers and the increase in enrollment in advanced classes suggest that the lack of achievement improvements arises in spite of what is generally viewed to be positive treatments. We also note that treatment results for 6th grade are stronger as they show peer differences of 0.37 to 0.46 standard deviations as well as larger differences in Vanguard class enrollment. These results are provided in Online Appendix Table 9. Below we investigate potential explanations for these findings, but first we turn to our analysis of GT magnet lotteries.

2.5. Estimates of the Impact of Attending a GT Magnet Using Randomized Lotteries

One reason the RD analysis does not show positive impacts from GT services on student outcomes may be that the qualification boundary is set low enough so that students who marginally qualify for GT services are not be able to take advantage of the purported benefits. Therefore, to examine other parts of the student quality distribution, in this section we present results using lotteries for the two GT over-subscribed magnet middle schools. Because the lottery is random, the comparison is across the entire distribution of those who apply. In fact, not only are the lottery students stronger than the marginal GT students in the RD sample, they are

³⁶ Reduced form and first stage results are provided in Online Appendix Table 8.

stronger than the average GT student in the district as shown in Table 2.1. A disadvantage, however, is that the lottery losers have a range of alternative experiences, although the bulk of them are in neighborhood GT programs. Nonetheless, students in the magnet GT schools with lotteries are shown to receive a more intense experience than students in other GT programs.

2.5.1. Data

Our lottery sample is derived from the set of 5th grade students determined to be eligible for GT in 2007-08 who apply for admission to one of the two middle schools with an over-subscribed GT magnet program.³⁷ We restrict our analysis to students who are observed to be enrolled in LUSD in 5th grade as these are the only students for whom we have pre-lottery characteristics. Also, this restriction reduces the likelihood of endogenous attrition as students who enter the lottery from outside LUSD would be more likely to leave if they lose the lottery, as many have previously attended private or charter schools. In addition, we drop students zoned to one of the schools with a regular program for students in the attendance zone.³⁸

While admission for non-zoned students is determined by lottery, our data does not directly provide the lottery ranking or outcome. Instead we identify whether a student is offered admission including those initially on a wait list.³⁹ In total the sample includes 542 students who participate in a lottery. Of these 394 are offered admission and 148 are not. By 7th grade 440 students including 331 winners (84%) and 109 losers (74%) remain in LUSD. The treatment

³⁷ The application process involves a single form where students may apply and rank to up to three of the eight magnet schools. Unfortunately our data only informs us of whether a person is offered a spot or wait listed; they do not have direct information on applications. Hence if a student is offered a spot at his or her first choice school we do not know if they applied for the other school. Nonetheless, we find no cases where a student is placed on the wait list for one of the lottery schools and offered a spot or waitlisted at the second while there are multiple instances whereby students are waitlisted at a lottery school and offered a slot at a non-lottery magnet. Hence, it appears that applying to both lottery magnets is very rare behavior in our data.

³⁸ The second school does not have zoned students, although it does include a program for students with severe physical disabilities. Students who are enrolled in this alternative program are not included in our lottery sample.

³⁹ Students with an older sibling in the school are exempt from the lottery, but as discussed in Section 2.3.2 above we believe the impact of this on our results is negligible.

received by the lottery losers varies, as they can attend GT classes in their neighborhood school, a charter, or a non-GT magnet school. Since there is some non-compliance with the lotteries we employ a 2SLS strategy that instruments GT magnet attendance with lottery outcomes.⁴⁰

2.5.2. Tests of Validity of Lottery Design

Table 2.6 presents the balancing tests for the lottery sample. The results strongly suggest that the lotteries for both magnet middle schools are conducted in a random way, as the ex-ante baseline (5th grade) sample has no significant coefficient on any of the twenty covariates we test.⁴¹ Further, using the ex-post estimation (7th grade) sample shows no significant differences between winners and losers except for math, which is significantly higher for winners at the 10% level. Although having one significant result out of twenty regressions can be spurious, it is nonetheless possible that this is due to differential attrition between lottery winners and losers. Indeed, when we estimate the impact of winning a lottery on attrition by 7th grade we find that lottery winners are 11 percentage points less likely to attrit (standard error of 0.04).

We thus use these results to inform our specification and analysis in three ways. First, as with the RD analysis, we present our results both with and without controls for lagged student scores as well as demographic characteristics. Second, we use a weighting procedure in the regressions that mimics the original lottery sample in order to correct for potential attrition bias. To do this we reweight the sample by the inverse of the predicted probabilities from a probit of attrition on 5th grade student characteristics.⁴² Third, we estimate bounds on the impact of GT using a procedure proposed by Engberg, et al. (2010). The procedure uses observable characteristics to estimate the proportion of the sample that includes students of various types

⁴⁰ By 7th grade 67% of lottery winners attend a magnet with a lottery while 17% attend another school and 16% leave the district. For lottery losers, 18% attend a lottery campus in 7th grade while 56% attend a different school and 26% leave the district.

⁴¹ Results for the 6th grade sample are similar as are results where standard errors are not clustered. These are provided in Online Appendix Tables 10 and 11.

⁴² Results of the probit regression are provided in Online Appendix Table 12.

including those who are at risk of leaving LUSD if they lose the lottery. A generalized method of moments (GMM) estimator is used to generate upper and lower bounds. The upper bound assumes students at risk of leaving due to losing have achievement equal to the mean of students who stay and comply with the lottery results, while the lower bound assumes these same students score at the 95th percentile of the outcome distribution for all staying participants.⁴³

2.5.3. Results for GT Magnet Programs

Two-stage least squares estimates of the impact on student achievement from attending one of the two magnet GT programs are shown in Table 2.7. Reduced-form estimates are provided in Online Appendix Table 13.⁴⁴ We provide both unweighted (rows 1 and 2) and inverse probability weighted (rows 3 and 4) estimates where the latter corrects for possible attrition bias. In rows (5) and (6) we provide upper and lower bounds that account for potential attrition bias using the Engberg, et al. (2010) methodology.

The results in Table 2.7 using our preferred specification of weighting with controls (row 4) suggest that, with the exception of science, which shows a 0.28 sd improvement, there is little impact of attending a GT magnet on achievement or attendance.⁴⁵ Due to the small sample sizes the estimates are somewhat imprecise, particularly using the inverse-probability weighted model. Even so, we note that the point estimates in row (4) for math, reading and social studies are negative and the estimate for language is effectively zero.⁴⁶ Hence, we believe these estimates

⁴³ That is, the upper bound assumes students at risk of leaving have only average scores, while the lower bound assumes they are in the upper tail. These assumptions are those suggested by Engberg, et al. (2010).

⁴⁴ The first stage is always significant at the 1% level with point estimates of 0.57 (standard error of 0.06) for unweighted and 0.47 (0.11) for weighted regressions. Detailed first-stage results are available upon request.

⁴⁵ Note that teacher manipulation is not a concern in this identification strategy, hence we can use the attendance results with confidence. Additionally, we do not provide discipline results as only 4% of students in the lottery sample have any disciplinary infractions in 7th grade.

⁴⁶ Results for 6th grade, provided in Online Appendix Table 14, show somewhat larger, albeit still insignificant in the preferred model, impacts for math and language and no impact for science. They also

provide strong evidence of a lack of positive impact of attending a magnet on achievement other than in science.⁴⁷ The bounding analysis in rows (5) and (6) confirm the results in row (4).⁴⁸ Once again we see little to suggest that there is any substantial positive impact on math, reading, language and social studies. For science, the lower bound does drop to zero which suggests that the positive result there may be due to attrition bias, but it nonetheless confirms that there is at least no negative impact on science scores.⁴⁹

In Table 2.8 we investigate the observable differences in treatment from attending a GT magnet relative to a neighborhood GT program.⁵⁰ The first five columns of the table show that using the weighted estimates, students who attend magnets gain peers, measured at the grade-course-teacher level, who score between 0.7 and 1.2 standard deviations higher than peers for neighborhood GT programs.⁵¹ Additionally, we find students who attend a GT magnet gain teachers whose value-added estimates are 0.09, 0.03 and 0.04 sd higher in math, English and Social Studies, respectively.⁵² Finally, in Online Appendix Table 18 we investigate whether there

show a significant negative impact on attendance of -0.6 percentage points (roughly one fewer day per year).

⁴⁷ Ceiling effects is a potentially even larger concern here than in the RD since the achievement levels of the lottery sample are higher. In Online Appendix Figures 14 - 18 we provide distribution plots of raw scores on 7th grade exams by lottery winners and losers. Although the mass of achievement is further to the right than in the RD sample, test score ceilings do not appear to be an issue as there appears to be substantial room for improvement.

⁴⁸ We do not provide bounding analyses for attendance as it performs poorly when the mean outcome is centered near a top-code as it tends to estimate outcomes to be above the top-code, which is the case in this sample since mean 5th grade attendance rates are 98.0 with a maximum of 100.

⁴⁹ Online Appendix Table 15 shows lottery results when we use attending a lottery magnet specifically as treatment (e.g. place non-lottery magnets in same category as neighborhood GT) and when we identify students who are taken off the wait list as losing the lottery. In both cases the results are similar to baseline except that science impacts become statistically insignificant.

⁵⁰ Estimates without controls and estimates for 6th grade are provided in Online Appendix Table 16 and are similar to those shown in Table 2.9.

⁵¹ In Online Appendix Table 17 we provide results assuming students are sorted randomly into sections, or are sorted by ability similar to Appendix Table 6 for the RD analysis. The results generally show similar levels of peer improvement where the difference in any case does not fall below 0.5 standard deviations.

⁵² As in the RD analysis, we also look at differences in course level. While there are no significant differences in 7th grade in the likelihood of enrolling in Vanguard courses, in 6th grade students who attend magnets are approximately 10 percentage points more likely to take Vanguard courses in math, English and social studies while they are 9 percentage points more likely to take Vanguard courses in science.

is variation in estimates by student types. The small sample limits how finely we can cut the data, nonetheless we find no discernable patterns across subpopulations. Hence it is clear that GT magnet students see improvements in their educational environment yet attain little improvement in achievement except in science. In the next section we discuss some potential explanations for the lack of positive impacts in both the RD and lottery analyses despite the apparent improvements.

2.6. Discussion

Given that GT students experience substantial treatments including better peers, more advanced courses (in the RD analysis), and higher quality teachers (in the lottery analysis), it is perplexing that we find little evidence of positive impacts on achievement. One possibility is that our achievement measure is not well suited to discerning improvements in gifted students. This would be particularly worrisome if we were to use a state accountability exam targeted towards low achieving students, but less of an issue with the Stanford Achievement Test. Indeed, we have already shown that there is little evidence of bunching near the maximum score (ceiling effects) in either the RD or lottery samples. While we cannot rule out that standardized tests poorly capture the effects of treatment, we note that the improvement in peers would be expected to generate higher achievement even if the exam is not well targeted to the curriculum.

Another potential explanation is that marginal students may suffer due to difficulty with more advanced material. In this view, the eligibility cut-off may be set at an inappropriate level as it leads the district to classify students as GT who are unable to deal with the advanced GT curriculum. While this explanation could be relevant for the RD results for marginal GT students, we demonstrate that the lottery sample includes higher achieving students for whom the advanced material would be more suitable. Further, our results are consistent with Abdulkadiroglu, Angrist

and Pathak (2011) and Fryer and Dobbie (2011) that elite schools have little impact on the achievement of marginal admits.

Given the strength with which peer effects have been found to operate in several different contexts, one would expect to find achievement improvements simply from the peer effects alone (Angrist and Lang, 2004; Hoxby and Weingarth, 2006; Lavy and Schlosser, 2011; Lavy, Paserman and Schlosser, 2008; Duflo, Dupas and Kremer, 2010; Imberman, Kugler and Sacerdote, forthcoming). Nonetheless, one possible reason for finding no impact of the differential GT resources is that the peer effects, in addition to the potential benefits found in the literature cited above, have potential costs as entering GT may reduce a student's relative ranking within the class (Davis, 1966). This could generate negative impacts through an invidious comparison model of peer effects where one's own performance falls with a reduction in one's position in the within-classroom achievement distribution (Hoxby and Weingarth, 2006). An alternative explanation with similar empirical predictions is that teachers may target the material in their classes to the median or higher achieving students (Duflo, Dupas, and Kremer, 2011).

Since we do not have direct evidence on student confidence, nor do we have direct evidence on how teachers target their classroom material, we empirically examine how the relative status changes for students in our two samples. Specifically, if student course grades and class rank changes based on admittance to a GT program or to a selective school, it is possible that the conditions for invidious comparison exist. Similarly, if course grades and rank change, it is possible that these changes may affect how well a student matches the ideal target student for a teacher. In Panel I of Table 2.9 we provide estimates of the impact of GT enrollment on course grades in the RD model, and of attending a GT magnet on grades in the lottery analysis. In both cases we find clear reductions in grades. For the RD sample GT student grades fall by a statistically significant 4 points out of 100 (3 points changes a grade from a B+ to a B, for example) in math compared to otherwise similar non-GT students, and by 2 to 3 points in other

subjects, although not statistically significant for 7th grade.⁵³ For the lottery analysis the grade reductions are even more dramatic for elite GT students, with drops of 7 points in math, 8 in science, and 4 in social studies using the inverse-probability weighted regressions.

In addition to the raw grades we consider how students' rankings within their peer groups differ by treatment status, as this provides a direct measure of how a student may perceive his or her relative achievement. We assume that students mostly compare themselves to students who take the same courses in the same grade. Thus, we rank students within each school-grade-course cell by their final course grades and convert these rankings to percentiles. Figure 2.7 demonstrates that the rankings based on 7th grade courses exhibit notable drops when students cross the GT eligibility threshold. In panel II.A of Table 2.9, we find even with student controls that marginal GT students have a relative rank 13 to 21 percentiles lower than marginal non-GT students in 7th grade. Panel II.B shows that attending a GT magnet in 7th grade generates a nearly 30 percentile ranking drop in all four of the courses examined.⁵⁴

To the extent that the negative estimates for grades and rank in the RD analysis reflects absolute changes in learning, this suggests that the more difficult course work could be ill suited to students at the eligibility margin. We would not, however, expect the lottery participants to be ill suited to more difficult coursework given their high positions in the GT achievement distribution. Additionally, we find little difference in curriculum as the likelihood of enrolling in a "Vanguard" class in 7th grade between lottery winners and losers is virtually identical. Thus, it seems possible that some portion of the grade effects reflect changes in relative rank independent of learning impacts.

⁵³We do not show reading as more advanced students do not take reading in 7th grade. Nonetheless, in the RD sample we also find a significant drop in reading grades in 6th grade, and among students who take reading in 7th grade of four points. For the lottery sample, however, only a handful of students take reading in 7th grade and hence the estimates are too imprecise to draw inferences.

⁵⁴Note that we limit the 6th grade results to the 2007-08 cohort only in this analysis so that the sample is comparable to that used in the 7th grade achievement analysis.

Nonetheless, these results only indicate that the necessary conditions for invidious comparison exist. To test this further, we estimate models in our RD analysis that interact GT enrollment with the student's 5th grade achievement in a subject. As previously discussed, since GT eligibility is determined by a number of factors, there is a substantial range of achievement in a given subject for students on the margin of eligibility. If IC is a key driver of our estimates, then we would expect to see students with higher achievement perform better in GT than students with lower achievement. We provide these results in the last row of Table 2.4 and find no statistically significant interaction estimates. However, this is a relatively weak test of the IC hypothesis and hence we cannot conclusively reject IC as an explanation.

2.7. Summary and Conclusion

In this chapter, we identify the impact of providing gifted and talented services on student achievement and behavior. We exploit a unique universal evaluation for GT in 5th grade which allows an RD analysis of achievement gains by 7th grade for students on the margin of eligibility. We also use lottery results from two elite magnet middle schools to assess achievement gains by 7th grade. Both samples offer a larger dispersion of student ability than is typical. In the RD dataset, the multiple criteria for eligibility results in a wide range of student achievement since students can be marginal in just a single dimension. For the elite school lottery sample, student applicants are on average very strong relative to even GT students in the district. The caveat in our two examples, however, is that the alternative to treatment varies. Marginal students not admitted to the GT program take “regular” classes, while students that lose the lottery receive GT services but in a less intensive atmosphere than that provided in the elite magnet schools.

Our analysis shows that both the RD and lottery samples meet the standard validity tests, with the main exception that lottery losers are more likely to leave the district. We correct for attrition in the lottery sample through inverse probability weighting for our estimates, and we also

generate coefficient bounds using the procedure proposed by Engberg, et al. (2010).

The RD results indicate that GT services generate little impact on achievement for students on the margin of qualifying. For the lottery analysis we also find little evidence of improvement in achievement or attendance with the significant exception of science. These results are on the surface surprising given that we find large improvements in peer achievement on the order of 0.3 standard deviations in the RD analysis, and 0.7 to 1.2 standard deviations in the lottery analysis. In addition, we find that students on the margin of GT eligibility enroll in more advanced classes, while students that gain admission to the premier GT magnet schools gain higher quality teachers. The estimates from these two samples and specifications are reduced forms, in that they do not differentiate among the many mechanisms by which student achievement might be impacted. Nonetheless, we are able to rule out many of the standard explanations for the lack of observed improvement in achievement.

One possible explanation for the surprising lack of measured improvement due to GT programs, despite the wide range of students we examine, may be shortcomings in using output from standardized multiple choice tests. Unfortunately, additional outcomes that may address this concern such as college attendance are not available yet for these cohorts. Hence, we leave this question to future research. Another possibility is that all of these students are strong, and will learn irrespective of formal educational inputs, but we are also unable to test this mechanism.

What we are able to measure, however, is that both raw course grades and students' relative rankings as measured by grades fall substantially in both the RD and lottery samples. While we do not find GT impacts that vary by prior achievement, operation of an IC model of peer effects remains a possibility. Irrespective of the exact mechanisms, however, our research is a broad contribution to the discussion of the limits to achievement gains that are possible through GT programs.

Figure 2.1: Gifted and Talented Matrix for GT Entry in 2008-09

STUDENT INFORMATION																																																																																		
Name: _____ Applying for Grade: _____																																																																																		
Date of Birth: _____ ID# _____ Ethnicity: _____																																																																																		
Zoned School: _____ Current School: _____																																																																																		
First Choice School: _____ Second Choice School: _____																																																																																		
ACHIEVEMENT TEST POINTS																																																																																		
<p style="text-align: center;">Stanford/Aprena 3 (within the last 12 months)</p> <table style="width: 100%;"> <tr> <th colspan="2" style="text-align: left;">Total Reading NPR</th> <th colspan="2" style="text-align: left;">Total Math NPR</th> </tr> <tr> <td>95-99 percentile</td> <td>12 points</td> <td>95-99 percentile</td> <td>12 points</td> </tr> <tr> <td>90-94 percentile</td> <td>10 points</td> <td>90-94 percentile</td> <td>10 points</td> </tr> <tr> <td>85-89 percentile</td> <td>8 points</td> <td>85-89 percentile</td> <td>8 points</td> </tr> <tr> <td>80-84 percentile</td> <td>6 points</td> <td>80-84 percentile</td> <td>6 points</td> </tr> <tr> <td>70-79 percentile</td> <td>4 points</td> <td>70-79 percentile</td> <td>4 points</td> </tr> </table> <p>Score: _____ Points: _____</p> <table style="width: 100%;"> <tr> <th colspan="2" style="text-align: left;">Total Science NPR</th> <th colspan="2" style="text-align: left;">Total Social Studies NPR</th> </tr> <tr> <td>95-99 percentile</td> <td>8 points</td> <td>95-99 percentile</td> <td>8 points</td> </tr> <tr> <td>90-94 percentile</td> <td>6 points</td> <td>90-94 percentile</td> <td>6 points</td> </tr> <tr> <td>85-89 percentile</td> <td>4 points</td> <td>85-89 percentile</td> <td>4 points</td> </tr> <tr> <td>80-84 percentile</td> <td>2 points</td> <td>80-84 percentile</td> <td>2 points</td> </tr> <tr> <td>70-79 percentile</td> <td>1 point</td> <td>70-79 percentile</td> <td>1 point</td> </tr> </table> <p>Score: _____ Points: _____</p> <p style="text-align: center;">Total Environment (Science/Social Studies) NPR (Grades K,1,2,3 only)</p> <table style="width: 100%;"> <tr> <td>95-99 percentile</td> <td>18 points</td> </tr> <tr> <td>90-94 percentile</td> <td>12 points</td> </tr> <tr> <td>85-89 percentile</td> <td>8 points</td> </tr> <tr> <td>80-84 percentile</td> <td>4 points</td> </tr> <tr> <td>70-79 percentile</td> <td>2 points</td> </tr> </table> <p>Score: _____ Points: _____</p>		Total Reading NPR		Total Math NPR		95-99 percentile	12 points	95-99 percentile	12 points	90-94 percentile	10 points	90-94 percentile	10 points	85-89 percentile	8 points	85-89 percentile	8 points	80-84 percentile	6 points	80-84 percentile	6 points	70-79 percentile	4 points	70-79 percentile	4 points	Total Science NPR		Total Social Studies NPR		95-99 percentile	8 points	95-99 percentile	8 points	90-94 percentile	6 points	90-94 percentile	6 points	85-89 percentile	4 points	85-89 percentile	4 points	80-84 percentile	2 points	80-84 percentile	2 points	70-79 percentile	1 point	70-79 percentile	1 point	95-99 percentile	18 points	90-94 percentile	12 points	85-89 percentile	8 points	80-84 percentile	4 points	70-79 percentile	2 points	<p style="text-align: center;">Aprena 2 (within the last 12 months)</p> <table style="width: 100%;"> <tr> <th colspan="2" style="text-align: left;">Total Reading NPR</th> </tr> <tr> <td>95-99 percentile</td> <td>20 points</td> </tr> <tr> <td>90-94 percentile</td> <td>14 points</td> </tr> <tr> <td>85-89 percentile</td> <td>9 points</td> </tr> <tr> <td>80-84 percentile</td> <td>6 points</td> </tr> <tr> <td>70-79 percentile</td> <td>4 points</td> </tr> </table> <p>Score: _____ Points: _____</p> <p style="text-align: center;">Total Math NPR</p> <table style="width: 100%;"> <tr> <td>95-99 percentile</td> <td>20 points</td> </tr> <tr> <td>90-94 percentile</td> <td>14 points</td> </tr> <tr> <td>85-89 percentile</td> <td>9 points</td> </tr> <tr> <td>80-84 percentile</td> <td>6 points</td> </tr> <tr> <td>70-79 percentile</td> <td>4 points</td> </tr> </table> <p>Score: _____ Points: _____</p>	Total Reading NPR		95-99 percentile	20 points	90-94 percentile	14 points	85-89 percentile	9 points	80-84 percentile	6 points	70-79 percentile	4 points	95-99 percentile	20 points	90-94 percentile	14 points	85-89 percentile	9 points	80-84 percentile	6 points	70-79 percentile	4 points
Total Reading NPR		Total Math NPR																																																																																
95-99 percentile	12 points	95-99 percentile	12 points																																																																															
90-94 percentile	10 points	90-94 percentile	10 points																																																																															
85-89 percentile	8 points	85-89 percentile	8 points																																																																															
80-84 percentile	6 points	80-84 percentile	6 points																																																																															
70-79 percentile	4 points	70-79 percentile	4 points																																																																															
Total Science NPR		Total Social Studies NPR																																																																																
95-99 percentile	8 points	95-99 percentile	8 points																																																																															
90-94 percentile	6 points	90-94 percentile	6 points																																																																															
85-89 percentile	4 points	85-89 percentile	4 points																																																																															
80-84 percentile	2 points	80-84 percentile	2 points																																																																															
70-79 percentile	1 point	70-79 percentile	1 point																																																																															
95-99 percentile	18 points																																																																																	
90-94 percentile	12 points																																																																																	
85-89 percentile	8 points																																																																																	
80-84 percentile	4 points																																																																																	
70-79 percentile	2 points																																																																																	
Total Reading NPR																																																																																		
95-99 percentile	20 points																																																																																	
90-94 percentile	14 points																																																																																	
85-89 percentile	9 points																																																																																	
80-84 percentile	6 points																																																																																	
70-79 percentile	4 points																																																																																	
95-99 percentile	20 points																																																																																	
90-94 percentile	14 points																																																																																	
85-89 percentile	9 points																																																																																	
80-84 percentile	6 points																																																																																	
70-79 percentile	4 points																																																																																	
ABILITY TEST POINTS																																																																																		
<p style="text-align: center;">Nagleri Nonverbal Abilities Test (NNAT) (current year's score)</p> <table style="width: 100%;"> <tr> <td>NAI 124-150</td> <td>30 points</td> </tr> <tr> <td>NAI 119-123</td> <td>25 points</td> </tr> <tr> <td>NAI 113-118</td> <td>20 points</td> </tr> <tr> <td>NAI 108-112</td> <td>15 points</td> </tr> <tr> <td>NAI 104-107</td> <td>10 points</td> </tr> <tr> <td>NAI 100-103</td> <td>5 points</td> </tr> </table> <p>Score: _____ Points: _____</p>			NAI 124-150	30 points	NAI 119-123	25 points	NAI 113-118	20 points	NAI 108-112	15 points	NAI 104-107	10 points	NAI 100-103	5 points																																																																				
NAI 124-150	30 points																																																																																	
NAI 119-123	25 points																																																																																	
NAI 113-118	20 points																																																																																	
NAI 108-112	15 points																																																																																	
NAI 104-107	10 points																																																																																	
NAI 100-103	5 points																																																																																	
REPORT CARD POINTS	TEACHER RECOMMENDATION	OBSTACLE POINTS																																																																																
<table style="width: 100%;"> <tr> <td>95-100</td> <td>Superior Progress</td> <td>20 points</td> </tr> <tr> <td>90-94</td> <td>Excellent Progress</td> <td>15 points</td> </tr> <tr> <td>85-89</td> <td>Very Good Progress</td> <td>10 points</td> </tr> <tr> <td>80-84</td> <td>Good Progress</td> <td>5 points</td> </tr> </table> <p>Matrix Score calculated using G/T Report Card Evaluation Rubric on page 2.</p> <p>Matrix Score: _____ Points: _____</p>	95-100	Superior Progress	20 points	90-94	Excellent Progress	15 points	85-89	Very Good Progress	10 points	80-84	Good Progress	5 points	<table style="width: 100%;"> <tr> <td>Score: 90-100</td> <td>10 points</td> </tr> <tr> <td>Score: 80-89</td> <td>8 points</td> </tr> <tr> <td>Score: 70-79</td> <td>6 points</td> </tr> <tr> <td>Score: 60-69</td> <td>4 points</td> </tr> </table> <p>Teacher Recommendation score calculated using G/T Identification Matrix on page 2.</p> <p>Score: _____ Points: _____</p>	Score: 90-100	10 points	Score: 80-89	8 points	Score: 70-79	6 points	Score: 60-69	4 points	<p>Check all appropriate boxes:</p> <p><input type="checkbox"/> Limited English Proficient</p> <p><input type="checkbox"/> Special Education/504</p> <p><input type="checkbox"/> Low SES</p> <p>(One or more = 5 points) Points: _____</p> <p>If Low SES Above + Minority(Hispanic or African American) = 3 additional points</p> <p style="text-align: right;">Total Points: _____</p>																																																												
95-100	Superior Progress	20 points																																																																																
90-94	Excellent Progress	15 points																																																																																
85-89	Very Good Progress	10 points																																																																																
80-84	Good Progress	5 points																																																																																
Score: 90-100	10 points																																																																																	
Score: 80-89	8 points																																																																																	
Score: 70-79	6 points																																																																																	
Score: 60-69	4 points																																																																																	
TOTAL	ADMISSIONS COMMITTEE																																																																																	
<p>TOTAL MATRIX SCORE: _____</p> <p>Score of 62 and above District Qualified.</p> <p>Score of 56 – 61 District Qualified if Stanford/Aprena scores equal 18 points and NNAT score equals 10 points.</p> <p>(Circle one)</p> <p>District Qualified - Not Qualified</p> <p>(Circle one)</p> <p>Vanguard Magnet - Vanguard Neighborhood</p>	<p>Meeting Date: _____ Date Information Sent to Parents: _____</p> <p>Committee Members: _____</p> <p style="text-align: center;">Campus G/T Coordinator – completed G/T Identification Matrix</p> <p>_____ G/T Committee Member – verified scores and points</p> <p>_____ VG Neighborhood Principal/ Designee or VG Magnet Advanced Academics Dept</p>																																																																																	

Figure 2.2: Gifted Status in 7th Grade
by 5th Grade Matrix Score

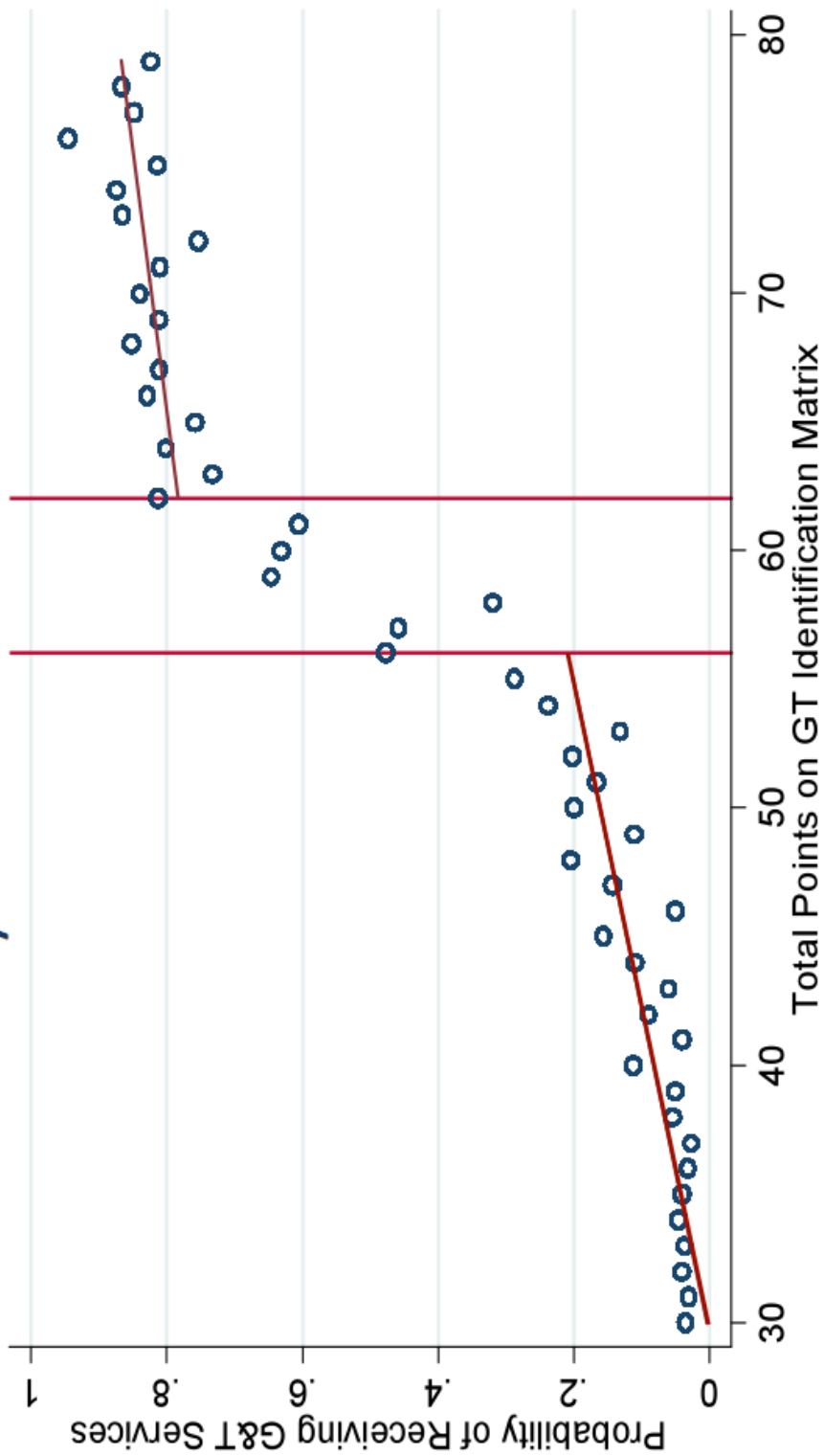


Figure 2.3: GT Qualification by Matrix Points

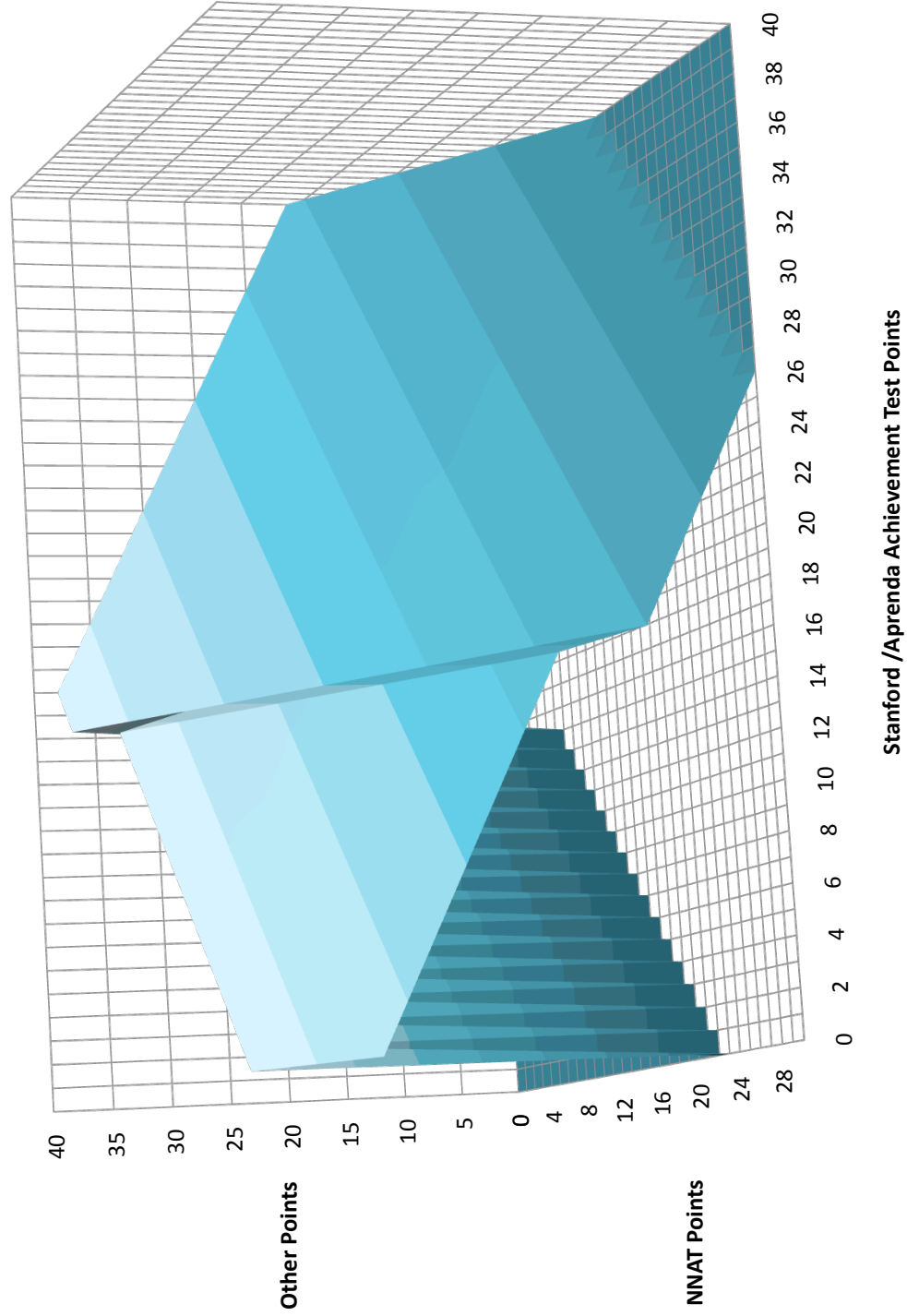


Figure 2.4: Gifted Status in 7th Grade by Distance to
Boundary Based on 5th Grade Matrix Points

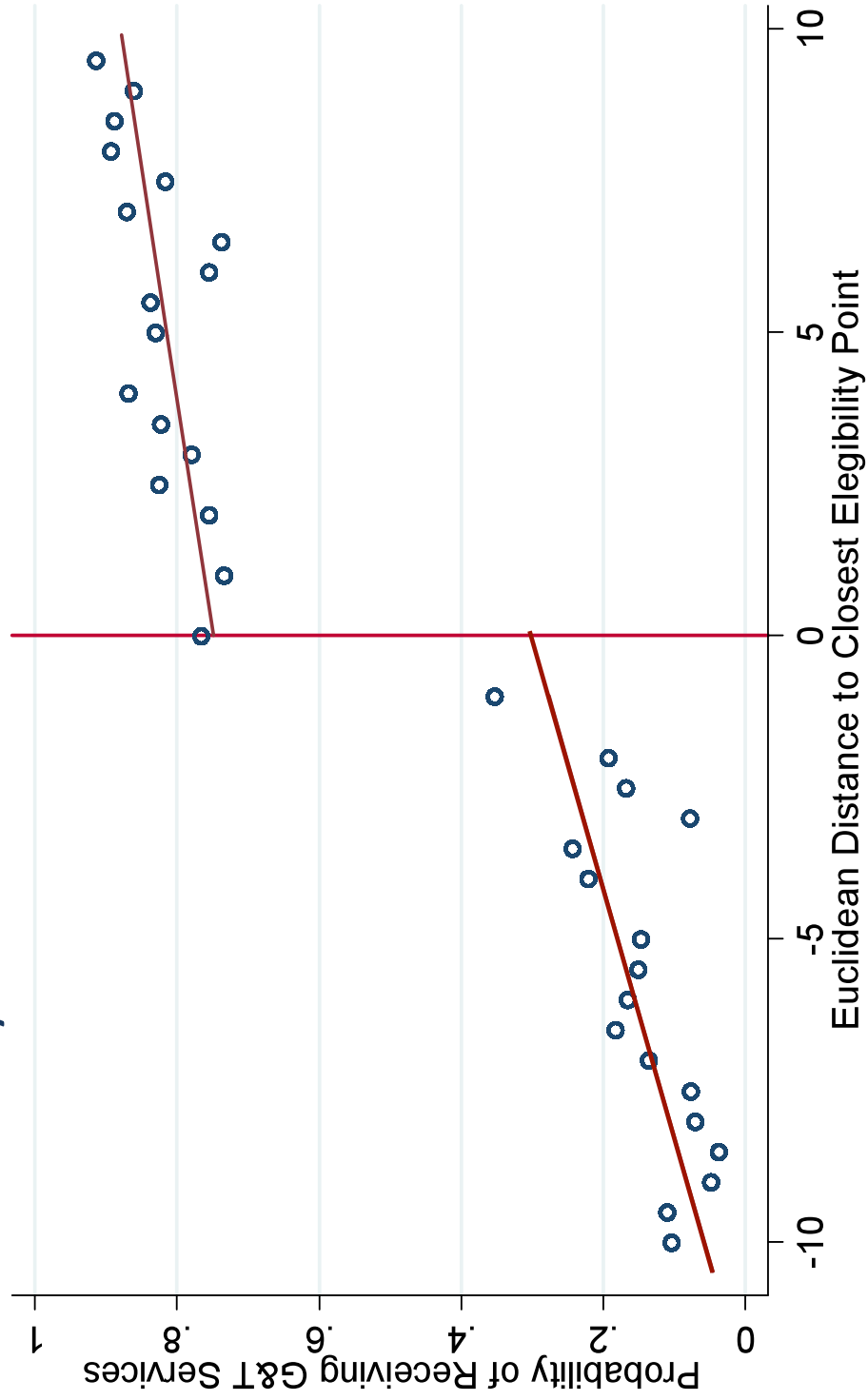


Figure 2.5: Distribution of Distances to Boundary

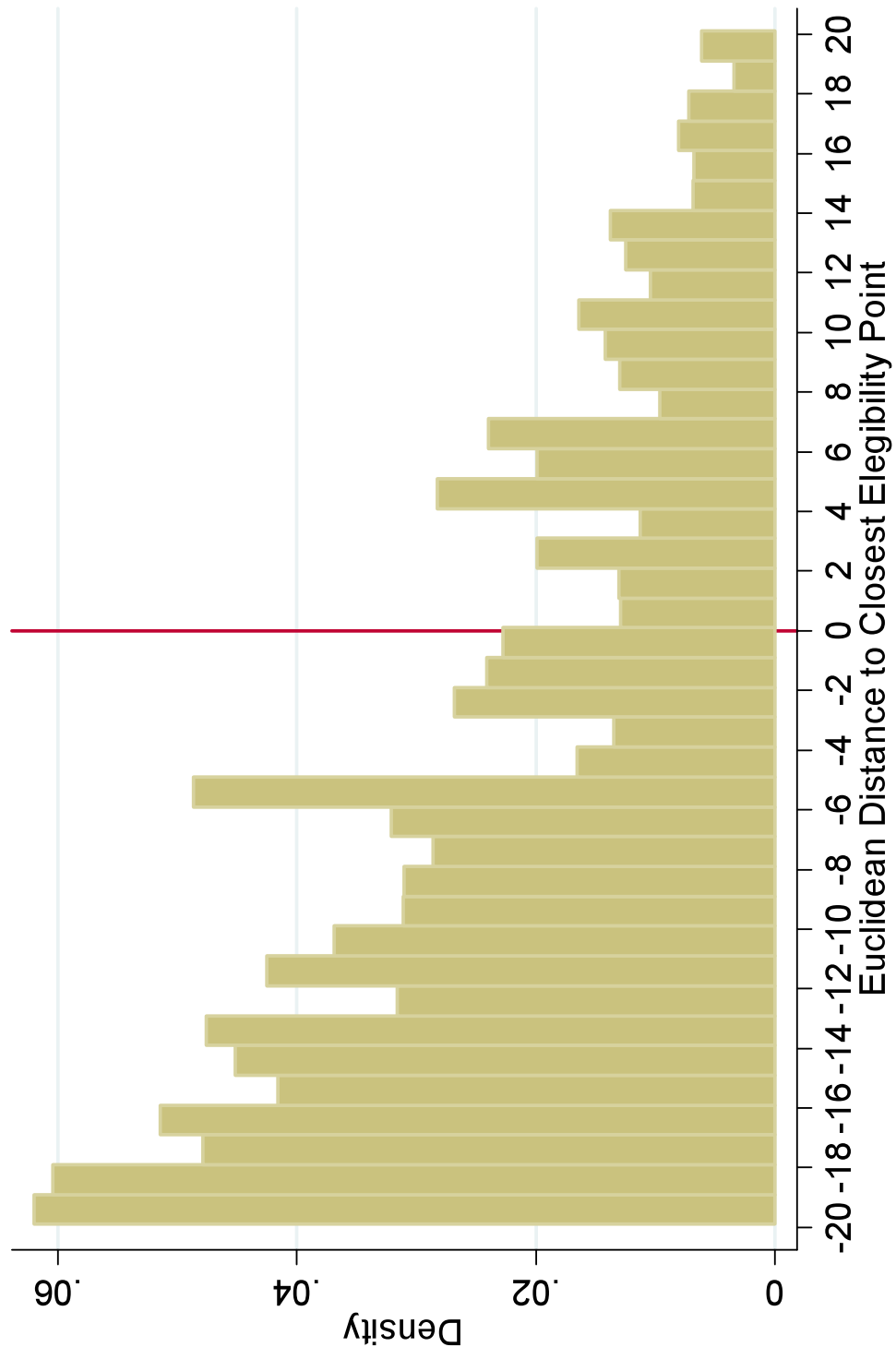
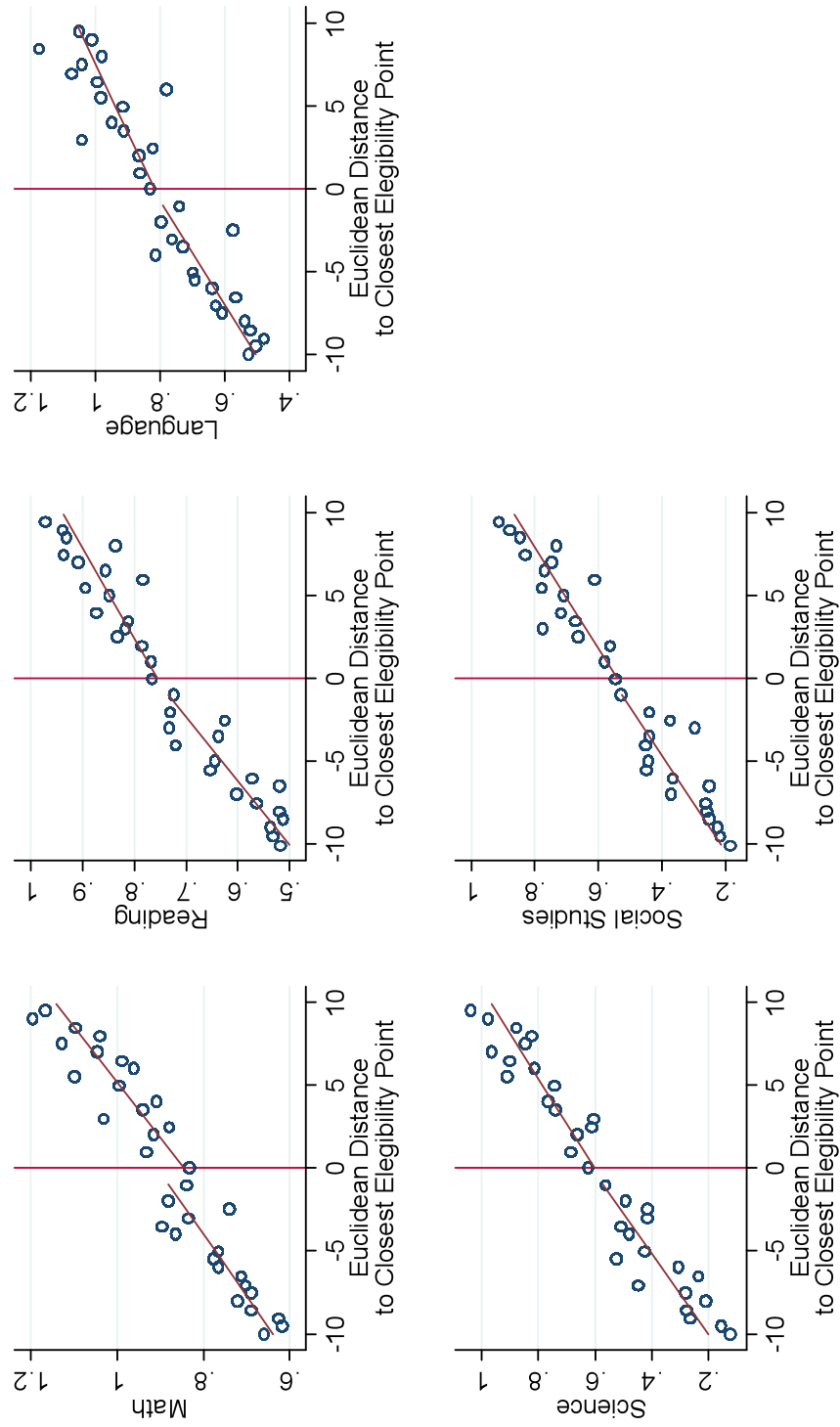
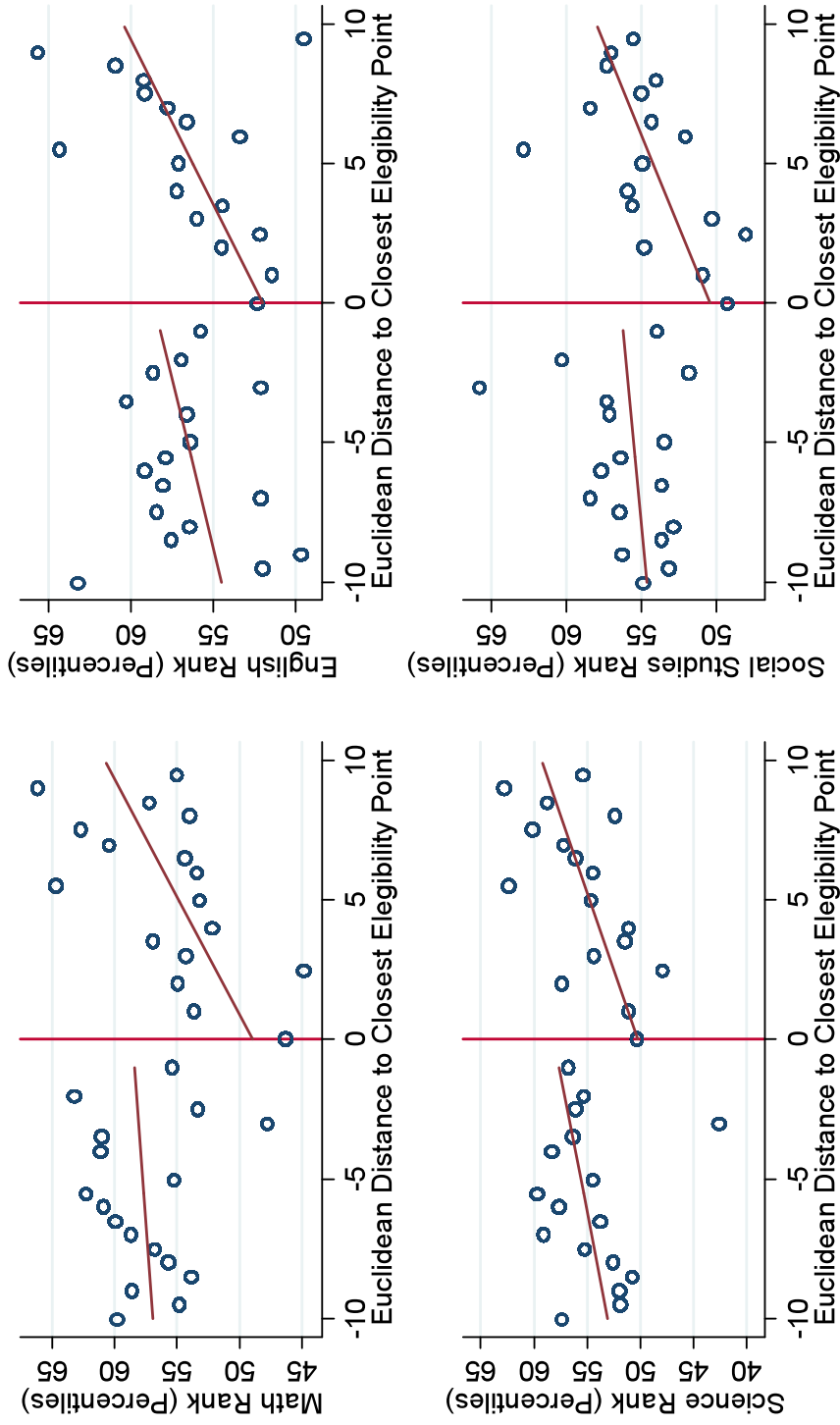


Figure 2.6: Reduced-Form Effects on Achievement 7th Grade
by Distance to Boundary



Achievement measured in standard deviation units within grade and year.

Figure 2.7: Rank in Course by Final Grade in
7th Grade by Distance to Boundary



Students with multiple courses in a subject are given the average rank over those courses.

Table 2.1. Characteristics of Students Evaluated for Middle School GT in 2007-08

	A. All 5th Grade Students			B. GT Magnet Lottery Sample		
	Gifted in 2009-10 (7th Grade)	Not Gifted in 2009-10	Not in Sample in 2009-10	In GT Magnet in 2009-10	Not in GT Magnet in 2009-10	Not in Sample in 2009-10
A. 5th Grade Characteristics						
Female	0.54 (0.50)	0.48 (0.50)	0.50 (0.50)	0.51 (0.50)	0.54 (0.50)	0.57 (0.50)
Economically Disadvantaged	0.59 (0.49)	0.89 (0.31)	0.81 (0.39)	0.24 (0.43)	0.41 (0.49)	0.17 (0.37)
LEP	0.23 (0.42)	0.37 (0.48)	0.28 (0.45)	0.02 (0.15)	0.06 (0.24)	0.04 (0.20)
Asian	0.11 (0.31)	0.02 (0.13)	0.03 (0.18)	0.28 (0.45)	0.16 (0.37)	0.19 (0.39)
Black	0.13 (0.34)	0.28 (0.45)	0.33 (0.47)	0.12 (0.32)	0.21 (0.41)	0.18 (0.38)
Hispanic	0.52 (0.50)	0.66 (0.47)	0.56 (0.50)	0.22 (0.41)	0.23 (0.42)	0.14 (0.35)
White	0.24 (0.43)	0.04 (0.19)	0.09 (0.28)	0.38 (0.49)	0.40 (0.49)	0.50 (0.50)
Gifted	0.68 (0.47)	0.06 (0.25)	0.15 (0.36)	0.85 (0.36)	0.85 (0.36)	0.83 (0.37)
Stanford Math	0.74 (0.59)	0.06 (0.39)	0.18 (0.47)	1.61 (0.79)	1.39 (0.71)	1.72 (1.03)
Stanford Reading	0.64 (0.41)	-0.02 (0.39)	0.11 (0.47)	1.72 (0.78)	1.60 (0.77)	1.83 (0.87)
Stanford Language	0.74 (0.59)	-0.16 (0.57)	0.01 (0.67)	1.61 (0.84)	1.48 (0.76)	1.83 (0.94)
Stanford Social Science	0.43 (0.68)	-0.61 (0.68)	-0.42 (0.80)	1.52 (0.86)	1.48 (0.84)	1.75 (0.91)
Stanford Science	0.50 (0.66)	-0.50 (0.65)	-0.30 (0.76)	1.47 (0.89)	1.36 (0.79)	1.61 (0.95)
Disciplinary Infractions	0.04 (0.26)	0.21 (0.73)	0.25 (0.87)	0.02 (0.15)	0.05 (0.24)	0.01 (0.10)
Attendance Rate	98.26 (2.35)	97.25 (4.52)	96.58 (4.95)	98.35 (2.00)	97.98 (2.34)	97.00 (3.75)
B. 7th Grade Outcomes						
Stanford Math	1.11 (0.45)	-0.40 (0.41)	-	1.70 (0.84)	1.53 (0.86)	-
Stanford Reading	0.95 (0.37)	-0.31 (0.38)	-	1.66 (0.66)	1.58 (0.72)	-
Stanford Language	1.08 (0.57)	0.17 (0.58)	-	1.59 (0.80)	1.44 (0.72)	-
Stanford Social Science	0.88 (0.64)	-0.09 (0.60)	-	1.70 (0.88)	1.51 (0.80)	-
Stanford Science	1.00 (0.79)	-0.18 (0.71)	-	1.72 (0.94)	1.36 (0.77)	-
Disciplinary Infractions	0.28 (1.11)	1.25 (2.61)	-	0.05 (0.24)	0.13 (0.86)	-
Attendance Rate	97.37 (3.19)	95.02 (6.13)	-	97.84 (2.52)	97.57 (3.16)	-
Observations	1,919	8,748	3,652	291	149	102

Standard deviations in parentheses. Achievement is measured in standard deviation units within grade and year across the district. Disciplinary infractions are the number of times a student is given a suspension or more severe punishment. Economically disadvantaged refers to students who qualify for free lunch, reduced-price lunch or another federal or state anti-poverty program.

Table 2.2. Reduced-Form Estimates of Discontinuities in Pre-Existing (5th Grade) Student Characteristics

	Black (1)	Hispanic (2)	Female (3)	LEP (4)	Gifted in 5th Grade (5)	Special Education (6)	Free / Reduced- Price Lunch (7)	Stanford - Math (8)	Stanford - Reading (9)	Stanford - Language Language (10)
Above GT Cutoff	-0.000 (0.029)	0.014 (0.038)	0.024 (0.042)	0.039 (0.040)	-0.050 (0.047)	0.005 (0.011)	0.049 (0.037)	-0.067*** (0.026)	0.006 (0.026)	0.006 (0.041)
Observations	2,650	2,650	2,650	2,650	2,650	2,650	2,650	2,637	2,638	2,636
	Stanford - Social Studies (11)	Stanford - Science (12)	# of Disciplinary Infractions (13)	Attendance Rate (%) (14)	Any Missing Matrix Data (15)	Teacher Score (16)	Teacher Points (17)	Enrolled (18)	Enrolled (Free/ Reduced-Price Lunch) (19)	Enrolled (Non- Free/ Reduced- Price Lunch) (20)
Above GT Cutoff	0.040 (0.049)	0.004 (0.042)	-0.001 (0.028)	-0.269 (0.190)	0.000 (0.008)	2.965 (2.715)	0.497 (0.321)	0.049 (0.030)	0.054 (0.037)	0.039 (0.053)
Observations	2,636	2,637	2,650	2,650	2,650	2,648	2,648	3,438	2,177	1,261

Achievement is measured in standard deviations of scale scores within grade and year. Disciplinary infractions are the number of infractions warranting a suspension or more severe punishment per year. Regressions include a linear smoother with a slope shift above the cutoff. The sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between -10 and 10. **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are robust to heteroskedasticity and clustered by 5th grade school. The estimation sample - students observed in LUSD two years after evaluation (7th grade) - is used in columns (1) to (17). Regressions using the full set of evaluated students provides similar results and is provided in the online appendix.

Table 2.3. Regression Discontinuity Estimates of Impact of Receiving GT Services

Model	Dependent Variable	Math	Reading	Language	Social Studies	Science	Disciplinary Infractions	Attendance Rate (%)
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Baseline								
OLS (RD Sample)	Enrolled in GT	0.157*** (0.044)	0.169*** (0.040)	0.254*** (0.049)	0.290*** (0.059)	0.312*** (0.068)	-0.328*** (0.091)	0.683*** (0.195)
Reduced Form	Above GT Cutoff	-0.061** (0.030)	-0.005 (0.029)	-0.004 (0.044)	-0.020 (0.038)	-0.011 (0.060)	-0.006 (0.120)	-0.691** (0.311)
2SLS - 1st Stage	Above GT Cutoff	0.440*** (0.057)	0.443*** (0.057)	0.442*** (0.058)	0.440*** (0.058)	0.440*** (0.057)	0.436*** (0.058)	0.438*** (0.058)
2SLS - 2nd Stage	Enrolled in GT	-0.138** (0.068)	-0.011 (0.065)	-0.008 (0.100)	-0.045 (0.085)	-0.025 (0.135)	-0.014 (0.276)	-1.578* (0.802)
Observations		2,612	2,614	2,612	2,610	2,612	2,653	2,652
B. With Individual Controls								
OLS (RD Sample)	Enrolled in GT	0.030 (0.023)	0.013 (0.017)	0.066** (0.026)	0.068*** (0.024)	0.069** (0.033)	-0.200** (0.089)	0.390** (0.193)
Reduced Form	Above GT Cutoff	-0.016 (0.022)	-0.001 (0.020)	0.005 (0.031)	-0.007 (0.031)	0.008 (0.048)	0.003 (0.112)	-0.502* (0.268)
2SLS - 1st Stage	Above GT Cutoff	0.465*** (0.060)	0.457*** (0.061)	0.457*** (0.061)	0.454*** (0.061)	0.456*** (0.061)	0.451*** (0.060)	0.456*** (0.060)
2SLS - 2nd Stage	Enrolled in GT	-0.035 (0.047)	-0.002 (0.044)	0.010 (0.068)	-0.016 (0.068)	0.017 (0.106)	0.007 (0.248)	-1.101* (0.653)
Observations		2,597	2,600	2,596	2,594	2,597	2,650	2,649
C. Using Synthetic Matrix Scores								
Reduced Form	Above GT Cutoff	-0.024 (0.028)	-0.028 (0.020)	-0.028 (0.039)	-0.054 (0.041)	0.002 (0.059)	0.088 (0.130)	0.346 (0.309)
2SLS - 1st Stage	Above GT Cutoff	0.229*** (0.038)	0.232*** (0.038)	0.230*** (0.039)	0.228*** (0.039)	0.229*** (0.038)	0.230*** (0.038)	0.229*** (0.038)
2SLS - 2nd Stage	Enrolled in GT	-0.106 (0.122)	-0.121 (0.085)	-0.120 (0.170)	-0.236 (0.188)	0.011 (0.256)	0.382 (0.568)	1.509 (1.328)
Observations		2,579	2,580	2,579	2,576	2,578	2,619	2,618

Achievement is measured in standard deviations of scale scores within grade and year. Disciplinary infractions are the number of infractions warranting a suspension or more severe punishment per year. Synthetic matrix scores replace matrix scores for students where a teacher recommendation could be pivotal (e.g. total points w/o the recommendation is fewer than 10 away from the relevant cutoff) with the predicted value from a regression of total points on all components excluding the teacher points. See text for details. Controls for race, gender, economic disadvantage, LEP, prior gifted status and lagged (5th grade) dependent variable included in panel B. All panels include a linear smoother with a slope shift above the cutoff. Sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between -10 and 10. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are robust to heteroskedasticity and clustered by 7th grade school.

Table 2.4. 2SLS Regression Discontinuity Estimates of Impact of Receiving GT Services
Specification Checks

	Stanford Achievement Test						Disciplinary Infractions (7)	Attendance Rate (%) (8)
	First Stage (1)	Math (2)	Reading (3)	Language (4)	Social Studies (5)	Science (6)		
(1) Quadratic Smoother	0.422*** (0.064)	0.120 (0.112)	0.007 (0.071)	0.246** (0.111)	0.146 (0.135)	0.305* (0.159)	-0.445 (0.505)	-0.565 (1.253)
Observations	2,609	2,597	2,600	2,596	2,594	2,597	2,650	2,649
(2) Cubic Smoother	0.371*** (0.103)	0.057 (0.238)	-0.029 (0.157)	0.276 (0.203)	-0.019 (0.244)	0.409 (0.332)	-0.617 (0.745)	-0.455 (2.036)
Observations	2,609	2,597	2,600	2,596	2,594	2,597	2,650	2,649
(3) Add Middle School Fixed Effects	0.460*** (0.057)	-0.014 (0.037)	0.007 (0.041)	0.041 (0.065)	0.009 (0.065)	0.023 (0.112)	0.067 (0.249)	-1.039* (0.600)
Observations	2,609	2,597	2,600	2,596	2,594	2,597	2,650	2,649
(4) Limited to Observations With No Missing Matrix Data	0.456*** (0.061)	-0.027 (0.048)	0.003 (0.044)	0.013 (0.067)	-0.004 (0.068)	0.029 (0.108)	0.068 (0.263)	-1.186* (0.684)
Observations	2,538	2,526	2,528	2,525	2,522	2,525	2,577	2,576
(5) Limit to Students Who Have 16 or More Stanford and 10 or More NNAT Points	0.892*** (0.317)	-0.024 (0.081)	0.003 (0.059)	0.135 (0.101)	0.003 (0.100)	-0.049 (0.128)	-0.433 (0.505)	-1.057 (0.923)
Observations	1,295	1,288	1,290	1,287	1,287	1,288	1,312	1,311
(6) Limit to Students Who Less than 16 Stanford or 10 NNAT Points	1.028** (0.510)	0.042 (0.099)	-0.069 (0.066)	0.082 (0.122)	-0.011 (0.121)	0.046 (0.169)	-0.008 (0.296)	-1.064 (0.941)
Observations	1,314	1,309	1,310	1,309	1,307	1,309	1,339	1,339
(7) Distance Between -4 & 4	0.391*** (0.085)	0.116 (0.167)	-0.097 (0.111)	0.132 (0.159)	-0.029 (0.170)	0.338 (0.246)	-0.762 (0.518)	-0.835 (1.647)
Observations	849	845	848	845	842	844	860	859
(8) Distance Between -8 & 8	0.462*** (0.056)	0.005 (0.058)	0.014 (0.046)	0.111 (0.072)	0.056 (0.080)	0.115 (0.103)	-0.162 (0.325)	-0.638 (0.758)
Observations	2,057	2,047	2,052	2,047	2,044	2,047	2,084	2,083
(9) Distance Between -12 & 12	0.472*** (0.055)	-0.009 (0.039)	0.018 (0.036)	-0.013 (0.057)	0.007 (0.063)	0.019 (0.086)	0.001 (0.209)	-0.823 (0.549)
Observations	3,178	3,162	3,163	3,158	3,158	3,160	3,222	3,220
(10) Distance Between -16 & 16	0.488*** (0.055)	-0.022 (0.035)	0.009 (0.030)	-0.015 (0.045)	-0.022 (0.061)	0.017 (0.077)	0.100 (0.179)	-0.438 (0.497)
Observations	3,756	3,735	3,736	3,731	3,729	3,733	3,806	3,804
(11) Local Linear Regressions with Rectangular Kernel	-	0.073 (0.117)	0.000 (0.072)	0.019 (0.186)	0.056 (0.080)	0.222 (0.177)	1.476 (1.002)	-0.434 (1.203)
Observations	-	1,075	1,078	708	2,044	1,074	429	1,092
Bandwidth for LLR (from Leave-One-Out Cross Validation)	-	5	5	3	8	5	2	5
(12) Interacting GT Impacts with Lagged Dependent Variable								
Enrolled in GT	-	-0.020 (0.048)	0.021 (0.047)	0.006 (0.067)	-0.018 (0.068)	0.019 (0.105)	0.032 (0.240)	13.509 (16.494)
Enrolled in GT * 5th Grade Dep Var	-	-0.029 (0.049)	-0.056 (0.047)	0.009 (0.030)	0.024 (0.041)	-0.054 (0.046)	-0.329 (0.334)	-0.149 (0.165)
Observations	-	2,597	2,600	2,596	2,594	2,597	2,650	2,649

Achievement is measured in standard deviations of scale scores within grade and year. Disciplinary infractions are the number of infractions warranting a suspension or more severe punishment per year. Controls for race, gender, economic disadvantage, LEP, prior gifted status and lagged (5th grade) dependent variable included and a linear smoother with a slope shift above the cutoff except where noted. Sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between -10 and 10. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are robust to heteroskedasticity and clustered by 7th grade school.

Table 2.5. 2SLS Estimates of Impacts of GT Services
Effects on Educational Environment and Student Choices

	Peer Math Scores in Math Classes (1)	Peer Reading Scores in Read/Eng Classes (2)	Peer Lang Scores in Read/Eng Classes (3)	Peer Soc Scores in Soc Classes (4)	Peer Science Scores in Science Classes (5)	# of Core Regular Classes (6)	# of Core Vanguard Classes (7)	Enrolled in Vanguard Math (8)	Enrolled in Vanguard English (9)
Enrolled in GT	0.348** (0.166)	0.287* (0.156)	0.311** (0.146)	0.235* (0.132)	0.272* (0.150)	-0.014 (0.267)	1.145* (0.624)	0.315* (0.158)	0.241 (0.171)
Observations	2,629	2,494	2,494	2,567	2,567	2,643	2,643	2,629	2,497

	Enrolled in Vanguard Social Science (10)	Enrolled in Vanguard Science (11)	Attends Zoned School (12)	Attends Non- Zoned GT Magnet Campus (13)	Attends Other Non-Zoned (14)	Math Teacher Fixed Effect (15)	Read/Eng Teacher Fixed Effect (16)	Science Teacher Fixed Effect (17)	Social Science Teacher Fixed Effect (18)
Enrolled in GT	0.282* (0.165)	0.282* (0.165)	-0.050 (0.109)	0.260** (0.109)	-0.210** (0.098)	-0.001 (0.025)	0.016 (0.010)	0.005 (0.014)	0.014 (0.013)
Observations	2,567	2,567	2,623	2,623	2,623	2,650	2,621	2,621	2,621

Achievement is measured in standard deviations of scale scores within grade and year. Teacher fixed effects are estimates from a student-level regression of achievement on lagged achievement, peer lagged achievement, race, gender, special education, LEP, at-risk status, teacher fixed-effects and school fixed-effects. Controls for race, gender, economic disadvantage, LEP, prior gifted status and lagged (5th grade) dependent variable included. Also includes a linear smoother with a slope shift above the cutoff. Peers are defined by teacher-course id-grade cells. The sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between -10 and 10. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are robust to heteroskedasticity and clustered by 7th grade school.

Table 2.6. Balancing Tests for GT Magnet Lotteries – Covariates Measured in 5th Grade

Sample	Stanford Achievement Test									
	Asian (1)	Black (2)	Hispanic (3)	White (4)	Econ Disadv (5)	Female (6)	At-Risk (7)	Special Education (8)	LEP (9)	Gifted (10)
Ex-Ante - Baseline Lottery	-0.030 (0.044)	0.030 (0.038)	0.041 (0.044)	-0.041 (0.050)	-0.035 (0.045)	-0.006 (0.047)	-0.011 (0.010)	-0.019 (0.017)	-0.033 (0.022)	-0.028 (0.035)
Observations	542	542	542	542	542	542	542	542	542	542
Ex-Post - Estimation Sample	-0.027 (0.048)	0.041 (0.038)	0.042 (0.055)	-0.057 (0.056)	-0.050 (0.059)	-0.001 (0.052)	-0.009 (0.011)	-0.015 (0.023)	-0.031 (0.027)	-0.024 (0.047)
Observations	437	437	437	437	437	437	437	437	437	437

Sample	Total Matrix									
	GT Magnet (11)	Points (12)	Math (13)	Reading (14)	Language (15)	Social Studies (16)	Science (17)	Attendance Rate (18)	Infractions (19)	Teacher Score (20)
Ex-Ante - Baseline Lottery	0.035 (0.030)	0.243 (0.926)	0.027 (0.069)	0.073 (0.063)	-0.034 (0.077)	0.053 (0.089)	0.010 (0.076)	-0.180 (0.201)	-0.022 (0.021)	0.029 (1.304)
Observations	542	542	540	541	539	540	539	542	542	536
Ex-Post - Estimation Sample	0.055 (0.045)	0.909 (1.173)	0.128* (0.074)	0.100 (0.075)	-0.059 (0.077)	0.063 (0.096)	0.090 (0.088)	-0.064 (0.230)	-0.022 (0.025)	-1.005 (1.471)
Observations	437	437	437	437	436	437	436	437	437	434

Achievement is measured in standard deviations of scale scores within grade and year. Disciplinary infractions are the number of infractions warranting a suspension or more severe punishment per year. Lotteries for two schools were conducted in 2007-08 hence regressions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student won the lottery. Robust standard errors clustered by 5th grade school in parentheses. Results without clustering are similar and provided in the online appendix.

Table 2.7. Effect of Attending a GT Magnet School Relative to a GT Neighborhood Program

	Model	Stanford Achievement Test				Attendance Rate (%)
		Math (1)	Reading (2)	Language (3)	Social Studies (4)	Science (5)
(1)	2SLS - Unweighted, No Controls	0.042 (0.178)	0.023 (0.103)	0.102 (0.065)	0.039 (0.083)	0.249** (0.114)
	Observations	437	438	436	437	440
(2)	2SLS - Unweighted, Controls	-0.100 (0.112)	-0.058 (0.105)	0.142* (0.081)	-0.032 (0.098)	0.208* (0.119)
	Observations	437	438	435	437	440
(3)	2SLS - Weighted, No Controls	-0.266 (0.291)	-0.130 (0.221)	-0.060 (0.148)	-0.120 (0.214)	0.243 (0.201)
	Observations	436	437	435	436	439
(4)	2SLS - Weighted, Controls	-0.224 (0.171)	-0.018 (0.172)	0.001 (0.114)	-0.036 (0.136)	0.281** (0.130)
	Observations	436	437	435	436	439
(5)	Engberg, Epple, Imbrogno, Sieg, Zimmer (2011) Bounds - Upper Bound	-0.019 (0.196)	-0.095 (0.157)	0.074 (0.162)	-0.064 (0.185)	0.344* (0.180)
	Observations	437	438	436	437	-
(6)	Engberg, Epple, Imbrogno, Sieg, Zimmer (2011) Bounds - Lower Bound	-0.353 (0.251)	-0.310 (0.192)	-0.207 (0.215)	-0.389 (0.249)	-0.013 (0.248)
	Observations	437	438	436	437	-

Achievement is measured in standard deviations of scale scores within grade and year. Lotteries for two schools were conducted in 2007-08 hence all regressions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student is enrolled in a GT magnet program in 7th grade. Robust standard errors clustered by 7th grade school in parentheses. Results without clustering are similar and provided in the online appendix. Controls include indicators during 5th grade for race, gender, special education, LEP, at-risk status, gifted, whether the student was enrolled in a GT magnet, and a lagged dependent variable. Weighted regressions are weighted by the inverse of the estimated probability of remaining in the data. See text for details. In order to avoid slow convergence due to a very small portion of the sample being in special education or LEP, we drop those controls from the bounding analysis. Additionally, we do not cluster the standard errors on the bounding analysis due to inability for the estimator to converge. Finally, we do not provide bounds for attendance due to poor performance with censored data. See paper for details.

Table 2.8. Treatments from Attending a GT Magnet School Relative to a GT Neighborhood Program

Model	Mean Peer Achievement (Std Deviations)				Average Teacher Effect			
	Math in Class	Reading in English Class	Language in English Class	Social Studies in Soc Class	Science in Science Class	Math	English/ Reading	Social Studies
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2SLS - Unweighted, Controls	1.066*** (0.145)	0.659*** (0.149)	0.579*** (0.120)	0.794*** (0.123)	0.524*** (0.122)	0.081*** (0.015)	0.032** (0.013)	0.031* (0.017)
Observations	440	436	436	439	439	440	440	440
2SLS - Weighted, Controls	1.164*** (0.179)	0.751*** (0.172)	0.686*** (0.143)	0.952*** (0.180)	0.659*** (0.166)	0.085*** (0.019)	0.032*** (0.011)	0.041** (0.019)
Observations	439	435	435	438	438	439	439	439

Achievement is measured in standard deviations of scale scores within grade and year. Average teacher effects use estimates for teacher dummy variables as the dependent variable. These estimated teacher effects come from a student-level regression of achievement on the teacher indicators, lagged achievement, peer lagged achievement, race, gender, special education, LEP, at-risk status, and school fixed-effects. Lotteries for two schools were conducted in 2007-08 hence all regressions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student is enrolled in a GT magnet program in 7th grade. Peers are defined by teacher-course id-grade cells. Robust standard errors clustered by 7th grade school in parentheses. Results without clustering are similar and provided in the online appendix. Weighted regressions are weighted by the inverse of the estimated probability of remaining in the data. See text for details. Controls include indicators during 5th grade for race, gender, special education, LEP, at-risk status, gifted, whether the student was enrolled in a GT magnet, and a lagged dependent variable. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 2.9. 2SLS Estimates of Impacts of GT on Course Grades and Rank (2007-08 Evaluation Cohort)

	I. Course Grades				II. Rank in Course (Percentiles)			
	Math (1)	English (2)	Social Studies (3)	Science (4)	Math (5)	English (6)	Social Studies (7)	Science (8)
A. Regression Discontinuity Analysis								
i. 7th Grade								
Enrolled in GT	-4.142** (1.616)	-2.621 (1.744)	-2.473 (1.645)	-1.501 (1.052)	-21.1*** (6.9)	-15.5** (7.1)	-17.1*** (5.8)	-13.2** (6.0)
Observations	2,643	2,510	2,581	2,602	2,643	2,510	2,581	2,602
ii. 6th Grade								
Enrolled in GT	-3.422*** (1.179)	-1.953 (1.491)	-2.931** (1.355)	-3.411** (1.442)	-17.9*** (6.2)	-16.9*** (6.5)	-22.6*** (7.0)	-22.9*** (6.9)
Observations	2,739	2,609	2,754	2,733	2,739	2,609	2,754	2,733
B. Lottery Analysis (7th Grade)								
Unweighted, Controls	-8.283*** (1.660)	-4.096** (1.561)	-4.062** (1.654)	-6.988*** (1.309)	-29.5*** (4.8)	-27.1*** (5.0)	-27.8*** (6.5)	-29.3*** (6.3)
Observations	440	437	439	439	440	437	439	439
Weighted, Controls	-7.311*** (1.847)	-2.719 (1.990)	-4.733** (1.733)	-8.121*** (2.297)	-30.7*** (5.4)	-30.4*** (7.7)	-33.8*** (6.8)	-36.1*** (8.9)
Observations	439	436	438	438	439	436	438	438

Rank is determined by rank-ordering the final grade in each course within school, grade and year, converted to percentiles. RD: Controls for race, gender, economic disadvantage, LEP, and prior gifted status are included along with a linear smoother with a slope shift above the cutoff. Sample is limited to students with Euclidean distances from qualifying via the GT qualification matrix of between -10 and 10. Standard errors are robust to heteroskedasticity and clustered by 7th grade school. Lottery: Lotteries for two schools were conducted in 2007-08 hence all regressions include indicators for lottery fixed effects. Coefficients are for an indicator for whether the student is enrolled in a GT magnet program in 7th grade. Peers are defined by teacher-course id-grade cells. Robust standard errors clustered by 7th grade school in parentheses. Results without clustering are similar and provided in the online appendix. Weighted regressions are weighted by the inverse of the estimated probability of remaining in the data. See text for details. Controls include indicators during 5th grade for race, gender, special education, LEP, at-risk status, gifted, and whether the student was enrolled in a GT magnet. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Chapter 3

Are Remittances in the Hands of Women more Effective? Evidence from Vietnam (with Adriana Kugler)¹

3.1. Introduction

International migration more than doubled in the past four decades, reaching 190 million in the late 2000s. Close to half of all international migrants come from the developing world and more than half of them are women, with 65% of all international migrants living in high-income countries.

Not surprisingly, international migration has been accompanied by a sharp rise in remittances, i.e., monetary transfers from migrants, back to their home countries. By 1997, international financial flows from remittances had surpassed overseas development assistance and by 2008 they were estimated to have reached 300 billion dollars. The evidence on the impact that remittances in terms of improving the lives of those left behind in their home country shows mostly positive effects. Some studies find that remittances contribute to the development of regions and households in the sending countries by improving health outcomes and increasing investment in education and capital (e.g., Cox and Ureta (2001), Yang (2008), Lopez-Cordova (2005), Gibson and McKenzie (2010)). However, a number of studies find negative impacts from the migration of household members on those remaining behind. For example, Gibson, McKenzie

¹ We thank Harry Holzer, Maurice Kugler, Ed Montgomery and Dean Yang for helpful comments.

and Stillman (2011) find that those left behind in the Pacific Islands are generally worse after the migration of other household members and McKenzie and Rapoport (2010) find that migration by household members in Mexico reduces educational attainment and attendance.

In this chapter we examine the impact of remittances on investments and the composition of spending in Vietnam and ask whether the impact of remittances depends on the gender of the person who receives the monetary transfer from the migrant. Under the unitary model of the household, household expenditure allocations should be independent of whether a man or woman in the household controls the money. On the other hand, if household decisions deviate from the unitary household model, then increases in the control of monetary resources within the household, say from increased remittances, will strengthen an individual's bargaining power and will change the allocation of expenditures. There is evidence from a number of countries that increased resources controlled by women at the time of marriage increases expenditure shares for education and health care (e.g., Quisumbing and Maluccio (2000), Thomas (1994), Hallman (2000) and Duflo (2003)). Here we ask whether increased control of remittances by women changes the allocation of expenditures and decisions within the household.

We use the Living Standards Measurement Surveys for Vietnam for 1992 and 1997 to examine the differential impact of remittances on household members when women control remittances. The empirical challenge to estimating the effects of remittances on households' outcomes comes from the fact that remittances and the fraction of remittances going to women may be endogenous. To address potential biases in the effects of remittances, we follow an instrumental variables strategy. Our instruments are regional migration rates coming from the 1992 survey and the interaction between this variable and the share of women in the household. Thus, our data sample for analysis come from the 1997 survey, while the instrument is derived from the earlier survey. OLS results show that the amount of remittances is associated with improved health, less adult employment, and increased business equipment. In addition, OLS results show that the greater fraction of remittances going to women, the greater educational

attainment of children in the household, the less employment for the young and adult, the less investment in businesses, the less household expenditure on food, and the higher the spending in health. However, our instrumental variables results show more limited effects. 2SLS results show a positive effect of remittances on household expenditure on education, but a negative effect on household expenditures on food items. Moreover, an increase in the share of remittances going to women increases household expenditures on health.

Our results, thus, suggest that not only do the amount of remittances affect investments, but that the gender of the receiver is also important in terms of how remittances affect households. The fact that whether a woman or a man receives the remittances matters in terms of households' decisions is inconsistent with unitary models of the household. To our knowledge, only two studies have examined the differential effects by the gender of remittance receivers. Guzman, Morrison and Sjoblom (2008) report results from simple OLS regressions and find that remittances going to female-headed household increase expenditures in health and education in Ghana. Gobel (2011) finds that female-headed households that receive remittances spend more on education and health and less on investment using household data from Ecuador. However, unlike our analysis, these studies only look at spending and do not examine other outcomes such as actual investments in human capital and equipment or on labor market outcomes.

3.2. Related Literature

The literature examining the impact of remittances has evolved from aggregate studies towards studies based on household and individual data. Adams and Page (2005) estimate the impact of remittances on poverty. Using data from 71 developing countries, they find that a rise in remittances reduces the share of those living in poverty, where remittances are instrumented with the distance from the remittance-sending area. Lopez-Cordova (2005) conducts a regional study for Mexico using the interaction between distance to the U.S. and historical migration as an instrument and finds that an increase in the fraction of households receiving remittances in a

municipality reduces infant mortality and child illiteracy and increases school attendance.

More recent studies have relied on individual-level data. A study by Cox-Edwards and Ureta (2003) uses the 1997 Annual household Survey from El Salvador to examine the impact of remittances on school attainment and controls for an indicator of whether the household received a remittance as a way to proxy for omitted variables. While the paper is not able to control for all omitted variables, the study finds that the probability of leaving school is lower when remittances increase. Acosta (2006) also uses data from El Salvador but instead uses matching techniques and finds that children in remittance receiving households have higher school attendance and lower employment than those in non-receiving households. Hanson and Woodruff (2003) and Borraz (2005) both use the 2000 Mexican Census and the interaction between the state migration rate and household characteristics and between the state migration rate and distance to the U.S. as instruments. Hanson and Woodruff (2003) find that remittances increase schooling overall, but Borraz (2005) finds that remittances only help to increase schooling in rural areas. Using GMM, Acosta et al. (2008) examine the impact of remittances using household data from 10 Latin American countries and find that remittances have negative but small effects on inequality and poverty. Finally, a study by Yang (2003) for the Philippines uses exchange rate shocks during the period of the Asian crisis and finds that an increase in remittances: raises school-related and investment-related expenditures, children's schooling, and the likelihood that a household enters an entrepreneurial activity. A recent paper by Gibson and McKenzie (2010) instead relies on matched difference-in-differences and finds that remittances from migrants to New Zealand increased income and consumption of more durable goods as well as child schooling in Tonga. However, another study looking at Pacific islanders by Gibson, McKenzie and Stillman (2011) finds that the absence of individuals allowed to migrate to New Zealand on the basis of a lottery has mostly negative impacts on those household members left behind. Similarly, McKenzie and Rapoport (2010) find lower schooling for young individuals in households with migrants in Mexico.

While the most reliable studies based on instrumental variables and matched difference-in-difference methods find positive effects of remittances on health, schooling and investments, none of these studies examines the differential impact of remittances going to women and men. A growing literature tests the unitary household model, i.e. testing whether household acts as one rational decision maker, in the context of developing countries. There is evidence that money in the hands of women has different effect on the outcomes of households' members compared to money in the hands of men. A number of studies find that unearned income in the hands of mothers increases education and health of children. Quisumbing and Maluccio (2003) find that women's assets at the time of marriage increase expenditure shares in education in Bangladesh and South Africa. Thomas (1990) finds that unearned income in the hands of mothers improves health of all children in Brazil, but Thomas (1994) finds that mothers' education has greater effects on daughters' height and fathers' education has a greater effect in sons' heights in Brazil and Ghana. Similarly, Hallman (2000) also finds that mothers' assets reduce daughters' morbidity while assets of fathers improve sons' health in Bangladesh. Duflo (2003) also documents gender asymmetries and finds that cash transfers to women have a positive impact on girls' weight for height and height for age measures but not on boys.

While many studies have shown evidence against the unitary model of the household in developing world few studies have examined the differential impact of remittances on households when women receive these remittances. Only Guzman, Morrison and Sjoblom (2008) and Gobel (2011) have examined the differential effects of remittances when received by a female vs. a male-headed household. Both of these studies look only at expenditures and find that remittances received by female-headed households raise the share of expenditures in health and education in Ghana and Ecuador. While Gobel tries to instrument for the amount of remittances, Guzman, Morrison and Sjoblom's (2008) data from Ghana does not allow them to use either matching or instrumental variables to deal with the endogeneity of remittances.

Furthermore, a recent review by Dean Yang in 2011 outlines exciting new research in the

area of migrant remittances besides that of the impact of remittances on receiving households. Particularly, Yang (2011) reviews literature on how much control migrants have over the money sent home or if they desire any control at all. This literature is similar to our study as they also examine the intra-household resource allocation. He reviews two studies. Ashraf, Aycinena, Martinez, and Yang (2011) conduct a randomized controlled trial among migrants from El Salvador living and working in the Washington, D.C. metro area. They find that migrants were more likely to open savings accounts when offered the option of greater control of the accounts. Chin, Karkoviata, and Wilcox (2010) study the effect of assistance of obtaining U.S. bank accounts on opening bank accounts and saving among Hispanic immigrants. They find that those assigned to the treatment experienced increased opening of U.S. bank accounts and higher savings in the United States and reduced remittances to Mexico. The effects are larger among those who report to have “no control” over how remittances are used in Mexico. These two papers study differential resource allocation between migrants and family members back home.

Here, we explore whether the effects of remittances differ with the gender of the receiver of the remittances in Vietnam. Contrary to the two previous studies that have looked at this question, we not only examine the impact on household expenditures but we also examine impacts on schooling, health, and labor market and investment outcomes. In addition, we address the endogeneity of remittances by providing instruments for both the amount of remittances as well as the share of remittances going to women.

3.3. Data

We use data from the Vietnamese Living Standards Surveys (VLSS) for the years 1992 and 1997. The Vietnamese Ministry of Planning and Investment along with the General Statistical Office (GSO) conducted the first VLSS between September 1992 and October 1993. GSO conducted the second survey between December 1997 and December 1998. These surveys were part of the Living Standards Measurement Study (LSMS) household surveys conducted in various

developing countries, with technical assistance from the World Bank. The surveys include information on the communities and the households. In our analysis, we focus on the household questionnaires, which collect information on demographic information characteristics, educational attainment, anthropometric measures, labor market activities, and place of residence. Most importantly, the surveys include detailed questions on the total amount of remittances from different sources, as well as the identity and location of the sender of the remittances and the identity of the receiver of the remittances. The data collector was asked first to list the names of the remitters, then correspondingly ask to write the down the ID code of the family member that received the money from each remitter.

The 1992 sample includes 4800 households and the 1997 sample includes the original 4800 households and an additional 1200 households, which were selected from the total sample of the 1995 Multi-Purpose Household Survey of the GSO. Our data sample of analysis comes from the 1997 sample. Specifically, we construct the variables using the answers to sample questionnaires by household members or representatives of households. Our education variables are the variable on the number of years of school a person has completed and whether the person currently attends school. Unfortunately, the questionnaires do not ask for the length of time the person takes to finish a certain grade. Only for education beyond high school do we know the time to completion. As a result, the former education variable assumes the person takes 1 year to finish each grade. Our labor market outcomes include whether the person was employed in the past 12 months and monthly salary of the job worked in the past 12 months. Our health variable is the person's body mass index (BMI), which is calculated as the person's weight in kilogram divided by the person's height in meter squared. Expenditures and total remittances a household received in the past 12 months are expressed in hundred thousand VN Dongs. We transform monetary values into real Dongs by deflating these with the 1997 regional and monthly CPIs.²

² As of April 30, 2012, according to www.xe.com, 1USD equals to 20,800VND.

Table 3.1 provides descriptive statistics of the variables for the various age sub-samples we consider in our analysis: young, adults, and older individuals. Individuals in remittance-receiving households in all age groups have more educated parents, are more likely to live in urban areas, and have higher schooling and attendance than those not receiving remittances. Of course, these differences should not be interpreted as causal. In fact, differences in parental schooling and urbanization may indicate self-selection into migration and remittance receipt and highlight the importance of controlling for these variables. Table 3.1 also shows that remittance-receiving households spend more, even though these households are on average smaller. In terms of the total of remittances, remittance-receiving households on average receive more than 40 hundred thousand VND in the past 12 months. This accounts for more than 20% of the annual household expenditures. Female household members receive roughly half of this total amount.

3.4. Empirical Framework

Remittances as extra income would relax liquidity constraints, allowing households to smooth consumption and invest in schooling, health and businesses. At the same time, having family members working elsewhere may disrupt family life that may bring about negative effect to outcomes of individual members and of households. For example, the absence of the mother or father may disrupt a child's school. Or the absence of the mother or father may also put pressure on the children to leave school and to earn money when there are no remittances. It is important to note, however, that in our 1997 survey, less than 2% of the remitters are wives/husbands of the receivers, while more than 48% of the remitters are children of the receivers. More than 20% of the remitters are sisters/brothers of the receivers.

The basic regression describing the relationship between total household remittances and individual outcomes is

$$Y_{ijk} = \beta R_{jk} + \rho S_{jk} + \Psi X_{ijk} + \Omega Z_{jk} + \Gamma \text{region}_k + \varepsilon_{ijk}, (1)$$

where Y_{ijk} is the outcome of interest of individual i in household j in region k . R_{jk} is the total amount

of remittances received by household j in region k , and S_{jk} is share of remittances going to women in the household j in region k . Vector \mathbf{X}_{ijk} contains the individual characteristics such as age and sex and the mother's and father's number of years of schooling. Vector \mathbf{Z}_{jk} contains household characteristics, which includes whether the household is in urban or rural area, whether the household is female-headed, and size of the household. Vector **region** $_k$ includes the regional-level controls, which are the proportion of the population living under the poverty line and the proportion of rural households in the region without allocated land. The World Bank estimated these two variables based on the 1992 survey (World Bank, 1999). However, we do not include such variables from the 1997 survey as migration itself can affect the regions' poverty and landless rates. Parameter ε_{ijk} is the individual error term, which may be correlated within households. Thus, we cluster standard errors at household level for individual-level regressions. We also estimate similar regressions for the household outcomes, but which do not control for individual characteristics.

The regressions above will provide us with relationships between total household remittances and the share of remittances received by women and the outcomes of interest, but the estimate on the total of remittances will not be causal. In particular, households receiving remittances may also be more likely to send children to school, to spend more money on healthcare and to invest in businesses. That is, observable and unobservable factors related with the receipt of remittances may also correlate with the outcomes of interest, which would bias the effects of remittances. We control for factors such as whether the household lives in an urban or rural area, region-level controls and the educational attainment of the mother and the father. However, unobservable factors such as motivation and drive may also be related to both the amount of remittances and investments. Likewise, the fraction of the remittances received by women in the household may be related to other factors. If the bargaining power of women in the household determines what fraction of the remittances women get, then this would capture exactly what we are interested in and there would be no bias. However, if other factors are

determining the fraction going to women and are also related to outcomes, then we would be getting biased results of these effects as well.

To establish a causal relationship between total household remittances and outcomes of interest, we rely on instrumental variables. Following Hanson and Woodruff (2003) and Hildebrandt and McKenzie (2005), one of our instruments is the migration rate from the 1992 survey. Specifically, it is calculated as follows:

$$1992_MigrationRate_k = \frac{Num_Migrants_k}{Num_People_k}, (2)$$

Where Num_People_k is the number of remittance senders in the region k and Num_People_k is the number of people in the region as they are calculated from the 1992 survey. There are 7 regions in Vietnam. From the survey questionnaires, migrants include internal migrants and those who work overseas as well as those who have settled permanently in another country. Specifically, in the 1992 survey, less than 20% of the remittance senders are living outside of Vietnam. Regarding the internal migration, the *doi moi* (renovation) program of the late 1980s has been the driving force behind the shift from organized to spontaneous migration in Vietnam (Niimi, Pham, and Reilly, 2009). According to Dang et al. (2003), there are three reasons for that change. The *doi moi* policy (1) rendered farmers less tied to the land from the de-collectivization in the agricultural sector, (2) the opening of the economy has allowed people to be less dependent on government subsidies, and (3) the increased flow of foreign direct investment into Vietnam has attracted migrants from various regions of the country to certain regions that have been the main recipients of these investments (e.g. the Southeast region). Additionally, it is suggestive in Small, Truong, and Vuong (2008) that the number of Vietnamese leaving Vietnam during different time periods depended on the region's connection to the political and legal circumstances. Thus, different regions have different migration rates. Furthermore, to satisfy the identifying assumption, the 1992 migration rate must not affect the outcomes through other factors, but only through the total remittances that households receive. We argue that by

controlling for the regions' characteristics in 1992, we hold constant the regions' persistent economic conditions that might have resulted in higher or lower migration out of the regions and differential individual and household outcomes.

In addition, the share of remittances received by women may also be endogenous, so we use the interaction between historical migration rate and the share of women in the household. The idea is that if there are more women in the household, remittances may just have to go to them. Overall, to identify parameters, β and ρ in equation (1), we use the 1992 migration rates and the interaction of this variable with the share of women in the household as the instrumental variables. Our first-stage regressions are thus:

$$\begin{aligned} R_{jk} &= \pi_0 1992_MigrationRate_k + \pi_1 1992_MigrationRate_k \times Share_Women_{jk} + \Sigma Z_{jk} + v_{jk} \\ S_{jk} &= \delta_0 1992_MigrationRate_k + \delta_1 1992_MigrationRate_k \times Share_Women_{jk} + \Lambda Z_{jk} + v_{jk} \end{aligned} \quad (3)$$

Then, we estimate the following 2SLS regression:

$$Y_{ijk} = \phi \hat{R}_{jk} + \theta \hat{S}_{jk} + \Phi X_{ijk} + \Delta Z_{jk} + \Pi region_k + \varepsilon_{ijk}, \quad (4)$$

where \hat{R} and \hat{S} are the predicted values from equations (3).

3.5. Results

Table 3.2 provides first-stage results for different samples used in OLS and IV regressions. For outcome variables that have the same samples of data, their first-stage results are the same and are not provided. All specifications contain whether household is in an urban area, household size, whether female heads the household, mother's and fathers' number of years of education, and 1992 region-level variables, which are the poverty rate and the rate of landlessness among rural households in the region. The individual-level regressions also contain information on the person's sex and age. Dependent variable in specifications on the two left columns of each panel is total of remittances; in the two right columns of each panel, it is fraction of remittances received by female members. F test gives the F-statistics of joint significance of the two IV

variables. While the first stage results in Panels A, B, C, E, and F appear to be less robust, in Panel D the instruments have a strong and significant relationship with total remittances and fraction received by women. The F statistics are greater than 12 for total remittances and greater than 40 for fraction of remittance received by women. For the household outcomes, the first-stage regression for total remittances shows that an increase of 0.02 i.e., one standard deviation in the 1992 migration rates, increases remittances by almost 400,000 VN Dong or about 10% increase in the average total remittances received in the sample. Thus, regions with historically high migration rates provide stable migration networks for the region and continue to have a positive association with future migration rates as reflected in the total household remittances. We also find that as the households with a greater fraction of women live in historically high migration rates, the share of women receiving remittances increases. Overall, for the household outcomes, we find the relationships between the endogenous variables and the instruments to be highly significant.

Tables 3.3-3.6 provide OLS and IV estimates of the impacts of remittances on various individual and household outcomes. Table 3.3 reports the regressions on educational outcomes, limiting to the sample of young people. Interestingly, OLS estimates reveal that total remittances have not associated with years of education or with attendance. When we control for the fraction of total household remittances received by women, we see that while total of remittances are negatively correlated with the number of years of schooling and with school attendance, the fraction is positively correlated with these two outcomes. An increase of 50% in the fraction of women receiving remittances is associated with about half an extra year of schooling and 0.09 higher probability of attending school. The following column in each panel displays the IV estimates of the impacts of remittances on education. We see that when we conduct 2SLS estimation, our standard errors increased greatly and that the IV estimates reveal insignificant relationship between remittances and the educational outcome of the young. Below the IV estimates are the F Statistics and p-value from the Hausman test for endogeneity of the two

variables of interest. For the outcome of the years of education, we can conclude that the variables are endogenous and that we should rely on the instruments. However, for the attendance outcome, we can conclude that IV estimates are inefficient.

The OLS estimates in Table 3.4 show the relationship between the body-mass index of young, adult and elderly members of the household and the total remittances and the share of remittances going to women. Interestingly, an increase in total remittances is associated with a higher BMI for all three groups, but the share of remittances received by women has little association with the BMI of these individuals. The IV estimates show that remittances have no significant impact on the BMI of the household members. Looking at the F-statistic and p-value from the Hausman test, the OLS and IV estimates are significantly different from each other for the adults, but not for the young and the elders.

Table 3.5 reports the relationship between total household remittances and employment of household members. Under OLS estimation, for young people, there is a positive but very weak relationship between total household remittances and the event that young people were employed in the past 12 months. An increase in remittances by 100,000 VN Dongs increases the likelihood of employment by 0.0006. On the other hand, an increase in the fraction of remittances received by females by 50% is associated with a lower probability of child labor of 0.08. However, there is no significant association between the two variables and the monthly salary. For the adults, total household remittances have a negative correlation with monthly salary in the past 12 months. Moreover, fraction of total household remittances going to women is negatively related with being employed. For the elders, total household remittances have a negative association with being employed. Once we instrument for total remittances and for the share received by women, we find little evidence that remittances have an impact on the employment of household members. Nor do we find that money receives by women has differential effect on employment outcomes. Examining the F-statistic and the p-value of the Hausman test, we can only reject the null hypothesis and conclude that the two variables of interest are endogenous in the monthly

salary outcome for the adults and not for the other samples of data.

Regarding OLS estimates of the impact of remittances on household outcomes, Table 3.6 reports the estimates on household expenditure and household non-farm enterprises. Total household remittances have positive association with profit from non-farm businesses, and the value of business equipment. On the other hand, the fraction of household remittances received by women has no significant association with business profit and a weak negative association with the value of business equipment. For households' expenditures, total household remittances have a negative association with the share of household expenditure that is food. Moreover, the share of remittances going to women has a positive association with the share of household expenditure that is health, but a negative association with the share of expenditure that is food.

The third column under each panel in Table 3.6 provides IV estimates of the effects of remittances on households' outcomes. We find that total remittances have positive and significant impact on households' business profit. Specifically, a 100,000 VND increase in total remittances increases this amount by more than 200,000 VND. We also find that total remittances have a significant and positive impact on households' share of total expenditure that is education. Particularly, a 100,000 VND increase in total remittances increases this share by almost 2 percent. At the same time, a 100,000 VND increase in total remittances decreases the share of household expenditure that is food by 1.6 percent. In terms of fraction of total remittances going to women, the share receives by household female members has a positive effect on the share of household expenditure that is health. Specifically, a 50 percent increase in this fraction increases the share by more than 5 percent. The IV estimates are bigger than the OLS estimates, suggesting that the OLS estimates were downwardly biased and that if anything those households where women receive a greater share of remittances also do worse in terms of household outcomes. Furthermore, the F-statistic and p-value suggest that the OLS and IV estimates differ significantly for the outcomes of business profit, education expenditure as a share of household expenditure, and health expenditure as a share of household expenditure.

In summary the IV results show that while total remittances are not associated with better health, lower adult employment, and higher salaries for those employed, they are positively with increased consumption in education and decreased consumption in food items by the households. The results also show that increasing women's control of remittance shifts household consumption toward health. Thus, control of monetary transfers from migrants by women matters in terms of increasing investment in health more than the actual amount remitted.

3.6. Conclusion

We use the Vietnamese Living Standards Surveys for 1992 and 1997 to examine whether women's control of remittances changes household investments and consumption patterns. Unlike the scarce literature that has tested the unitary model of the household exploiting financial resources coming from women, we use instrumental variable techniques to eliminate potential biases due to the endogeneity of total remittances and of the share of remittances received by women. Our IV results are bigger than the OLS results suggesting that those receiving more remittances overall and those that send more money to women in the households are less inclined to invest and spend. While some of our first stage results are not robust, the evidence is strong in showing that increasing overall remittances has positive effects in terms of increasing household consumption in education and household business profit.

More importantly, our results show that while increased control of remittances by women has little impact children outcomes, it shifts households' expenditure toward health. This result suggest that when women are given greater bargaining power they will sway decisions towards investment in household health and education than expenditure on food items. Increasing women's decision-making within the household is important in terms of raising human capital for family members.

Table 3.1. Basic Descriptive Statistics of Remittance-Receiving and Non-Receiving Households in the 1997-1998 Survey

	Young (age ≥ 5 & age < 18)		Adults (18 ≤ age < 65)		Elders (age ≥ 65)	
	Remittance-Receiving Households	Non-Receiving Households	Remittance-Receiving Households	Non-Receiving Households	Remittance-Receiving Households	Non-Receiving Households
Individual-level Variables						
Female	0.490 (0.500)	0.487 (0.500)	0.548 (0.498)	0.517 (0.500)	0.592 (0.492)	0.583 (0.493)
Age	11.033 (3.709)	11.268 (3.651)	37.669 (13.540)	36.227 (12.890)	73.449 (6.121)	73.173 (6.181)
Number of Years of Schooling of Father	8.605 (3.738)	6.687 (4.372)	4.989 (4.105)	3.808 (3.617)	1.645 (2.647)	1.532 (2.643)
Number of Years of Schooling of Mother	6.860 (3.412)	5.017 (4.155)	2.418 (2.963)	1.854 (2.565)	0.294 (1.059)	0.347 (1.084)
Currently Attending School	0.842 (0.365)	0.789 (0.408)	0.045 (0.208)	0.035 (0.184)	0.000 (0.000)	0.000 (0.000)
Number of Years of Schooling	4.874 (3.455)	4.587 (3.220)	7.955 (4.016)	6.863 (3.909)	3.024 (3.488)	2.558 (3.165)
Employed in the Past 12 Months	0.225 (0.418)	0.312 (0.463)	0.834 (0.372)	0.910 (0.287)	0.380 (0.486)	0.382 (0.486)
Monthly Salary of the Job Worked in the Past 12 Months (zero if self-employed)	0.281 (0.961)	0.193 (0.854)	1.546 (3.154)	0.948 (2.533)	0.112 (0.647)	0.097 (0.549)
BMI	15.785 (2.389)	15.824 (2.385)	20.219 (2.690)	19.981 (2.450)	19.251 (3.023)	18.752 (2.712)
Regional Variables						
1992 Migration Rate	0.070 (0.018)	0.065 (0.022)	0.074 (0.017)	0.067 (0.021)	0.073 (0.017)	0.067 (0.021)
1992 Poverty Rate	0.560 (0.166)	0.587 (0.156)	0.541 (0.164)	0.570 (0.157)	0.564 (0.158)	0.581 (0.150)
1992 Rate of Landlessness among Rural Households	0.106 (0.076)	0.093 (0.073)	0.113 (0.078)	0.100 (0.074)	0.100 (0.076)	0.094 (0.072)

	Youngs (age ≥ 5 & age < 18)		Adults (18 ≤ age ≤ 65)		Elders (age > 65)	
	Remittance-Receiving Households	Non-Receiving Households	Remittance-Receiving Households	Non-Receiving Households	Remittance-Receiving Households	Non-Receiving Households
Household-level Variables						
Household Total of Remittances	44.760 (106.311)	0.000 (0.000)	48.539 (106.766)	0.000 (0.000)	42.942 (92.448)	0.000 (0.000)
Fraction of Remittances Received by Females	0.482 (0.483)	0.000 (0.000)	0.496 (0.476)	0.000 (0.000)	0.444 (0.469)	0.000 (0.000)
Total of Non-farm's Businesses' Profit	71.806 (200.445)	54.152 (149.950)	75.171 (182.220)	63.221 (177.428)	43.414 (120.797)	46.129 (148.058)
Total Value of Non-farm Businesses' Equipment	69.750 (381.369)	57.543 (344.940)	63.482 (314.222)	71.513 (472.451)	46.999 (247.519)	68.162 (599.909)
Female-headed Household	0.281 (0.450)	0.159 (0.366)	0.325 (0.469)	0.218 (0.413)	0.335 (0.472)	0.245 (0.430)
Urban	0.342 (0.474)	0.182 (0.386)	0.461 (0.499)	0.261 (0.439)	0.381 (0.486)	0.255 (0.436)
Size of Household	5.676 (1.910)	5.940 (1.866)	5.062 (2.107)	5.516 (2.027)	3.896 (2.194)	5.286 (2.250)
Share of Education Expenditure	0.068 (0.057)	0.059 (0.057)	0.055 (0.072)	0.049 (0.064)	0.024 (0.046)	0.032 (0.051)
Share of Health Expenditure	0.061 (0.078)	0.046 (0.057)	0.066 (0.088)	0.048 (0.060)	0.096 (0.105)	0.063 (0.071)
Share of Food Expenditure	0.532 (0.142)	0.588 (0.138)	0.514 (0.145)	0.574 (0.144)	0.539 (0.146)	0.586 (0.146)
Total Household Expenditure	173.663 (138.398)	141.007 (105.800)	185.557 (149.694)	150.120 (114.958)	144.385 (164.048)	135.528 (115.235)
Max N	7,263	1,816	3,550	11,820	821	1,009

NOTES: "Number of Year of Schooling" refers to the number of grade (gradeschool) and years of schooling the person completed. Monetary variables are in 100,000 VND and are inflated by regional and monthly indices and under the time frame of the past 12 months. Standard errors are in parentheses.

Table 3.2. First Stage Results

	A. Education				B. BMI					
	Two Educational Outcomes		(3)	(4)	Young		Adults		Elders	
	(1)	(2)			(5)	(6)	(7)	(8)	(9)	(10)
1992 Migration Rate	227.1140 (446.8470)	-1.2021 (2.2158)			313.5346 (411.3629)	-1.3091 (2.1779)	156.8304* (82.1778)	0.3568 (0.3547)	719.9637 (441.3871)	1.1095 (1.4875)
1992 Migration Rate X Fraction of Female in the Household	-1,013.2992 (1,010.8484)	5.4216* (3.0108)			-894.2381 (1,064.8077)	5.2335* (3.0976)	68.3121 (105.9511)	1.8395*** (0.4450)	-286.2774 (302.2582)	1.9541 (1.7983)
F-Test	0.5900	1.7600			0.3600	1.5000	6.4600	16.9600	2.7200	1.8900
p-value	0.5565	0.1758			0.7005	0.2275	0.0016	0.0000	0.0674	0.1524
N	183	183			174	174	6,308	6,308	360	360
C. Employment: Young					D. Household's Outcomes					
	Employed in the Past 12 Months		Monthly Salary from job in the Past 12 Months		Other Household's Outcomes		Value of Business Equipment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1992 Migration Rate	153.8887 (489.9729)	-1.1181 (2.3613)	929.2646 (928.8807)	-2.4152 (3.7345)	198.4789*** (53.2580)	-0.0967 (0.3169)	201.2300*** (52.6197)	-0.1415 (0.3190)		
1992 Migration Rate X Fraction of Female in the Household	-972.2024 (1,008.3995)	5.4984* (3.0347)	-2,093.7656 (2,073.9760)	14.0977* (7.1382)	13.4816 (58.8696)	2.9237*** (0.3470)	-9.4397 (53.0151)	2.9529*** (0.3489)		
F-Test	0.6500	1.8400	0.5200	2.6600	14.24	47.6000	12.9800	47.2500		
p-value	0.5250	0.1626	0.5970	0.0808	0.0000	0.0000	0.0000	0.0000		
N	177	177	52	52	5,998	5,998	5,851	5,851		
E. Employment: Adults					F. Employment: Elders					
1992 Migration Rate	158.4044** (80.6483)	0.4053 (0.3551)	149.2732* (84.6333)	0.4569 (0.3557)	706.8372 (436.8953)	1.4451 (1.4712)	125.2725* (66.8185)	1.3944 (1.8476)		
1992 Migration Rate X Fraction of Female in the Household	62.9458 (103.9380)	1.8706*** (0.4414)	53.0913 (109.7639)	1.6001*** (0.4452)	-258.2537 (293.8646)	1.4054 (1.8128)	92.7012 (68.5262)	2.7008 (2.4319)		
F-Test	6.5800	18.3300	5.6900	14.9500	2.8600	1.7400	7.0700	2.6200		
p-value	0.0014	0.0000	0.0034	0.0000	0.0590	0.1764	0.0011	0.0755		
N	6,437	6,437	5,889	5,889	371	371	202	202		

Notes: First-stage results are presented for all outcomes when the sample of data varies. Robust standard errors are reported in parentheses. All specifications contain whether household is in an urban area, household size, whether female heads the household, mother's and fathers' number of years of education, and 1992 region-level variables, which are the poverty rate and the rate of landlessness among rural households in the region. The individual-level regressions also contain information on the person's sex and age. Dependent variable in specifications on the two left columns of each panel is Total of remittances; in the two right columns of each panel, it is Fraction of remittances received by female members. F test gives the F-statistics of joint significance of the two IV variables. *** p<0.01, ** p<0.05, * p<0.1

Table 3.3. OLS and IV Estimates of the Impact of Remittances on Education

	Years of Education			Attendance		
	OLS		IV	OLS		IV
	(1)	(2)	(3)	(4)	(5)	(6)
Total of Remittances	-0.0011 (0.0011)	-0.0029** (0.0013)	3.0851 (701.9062)	-0.0003 (0.0002)	-0.0006*** (0.0002)	0.5825 (131.9304)
Fraction received by Female Members		1.0062** (0.3857)	583.7288 (130,952.5904)		0.1897** (0.0762)	109.4436 (24,613.7277)
F-Stats from Hausman Test			5.04			0.6800
p-value from Hausman Test			0.0077			0.5084
N	183	183	183	183	183	183

Notes: Robust standard errors are clustered at household level and reported in parentheses. All specifications contain age, sex, whether household is in an urban area, household size, whether female heads the household, mother's and fathers' number of years of education, and 1992 region-level variables, which are the poverty rate and the rate of landlessness among rural households in the region. The instruments are migration rates from Vietnam's 7 regions coming from the 1992 survey and the interaction of this variable and the fraction of female in the households. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.4. OLS and IV Estimates of the Impact of Remittances on Health

	BMI								
	Young		Adults				Elders		
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total of Remittances	0.0039** (0.0019)	0.0039* (0.0022)	0.1078 (0.5780)	0.0013* (0.0007)	0.0014* (0.0008)	-0.0615 (0.0466)	0.0055*** (0.0015)	0.0053*** (0.0015)	-0.0360 (0.0347)
Fraction received by Female Members		0.0232 (0.4687)	20.3363 (95.6882)		-0.1169 (0.1222)	2.1609 (5.2223)		0.3321 (0.5069)	4.6998 (5.9431)
F-Stats from Hausman Test			1.0100			8.6700			2.4900
p-value from Hausman Test			0.3658			0.0002			0.0847
N	174	174	174	6,308	6,308	6,308	360	360	360

Notes: Robust standard errors are clustered at household level and reported in parentheses. All specifications contain age, sex, whether household is in an urban area, household size, whether female heads the household, mother's and fathers' number of years of education, and 1992 region-level variables, which are the poverty rate and the rate of landlessness among rural households in the region. The instruments are migration rates from Vietnam's 7 regions coming from the 1992 survey and the interaction of this variable and the fraction of female in the households. *** p<0.01, ** p<0.05, * p<0.1

Table 3.5. OLS and IV Estimates of the Impact of Remittances on Employment

	A. Young					
	Employed in the Past 12 Months			Monthly Salary from job in the Past 12 Months		
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Total of Remittances	0.0003 (0.0002)	0.0006** (0.0002)	-0.0279 (0.3883)	-0.0015** (0.0007)	-0.0009 (0.0006)	0.0029 (0.0099)
Fraction received by Female Members		-0.1548* (0.0932)	-4.8762 (71.3824)		-0.3464 (0.2204)	0.0669 (1.1134)
F-Stats from Hausman Test			0.3900			0.0600
p-value from Hausman Test			0.6791			0.9381
N	177	177	177	52	52	52
	B. Adults					
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Total of Remittances	-0.0001 (0.0001)	-0.0001 (0.0001)	0.0000 (0.0021)	-0.0010* (0.0005)	-0.0011** (0.0005)	0.0187 (0.0290)
Fraction received by Female Members		-0.0472*** (0.0134)	0.0382 (0.2213)		0.0788 (0.1166)	2.0587 (3.5304)
F-Stats from Hausman Test			0.4200			8.3000
p-value from Hausman Test			0.6593			0.0003
N	6,437	6,437	6,437	5,889	5,889	5,889
	C. Elders					
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
	OLS (1)	IV (2)	IV (3)	OLS (4)	IV (5)	IV (6)
Total of Remittances	-0.0010** (0.0004)	-0.0010** (0.0004)	-0.0017 (0.0045)	0.0017 (0.0030)	0.0024 (0.0028)	0.0924 (0.1919)
Fraction received by Female Members		0.0097 (0.0788)	0.7290 (1.1491)		-0.0876 (0.1188)	-5.6488 (10.9479)
F-Stats from Hausman Test			0.3100			1.7500
p-value from Hausman Test			0.7366			0.1775
N	371	371	371	202	202	202

Notes: Robust standard errors are clustered at household level and reported in parentheses. All specifications contain age, sex, whether household is in an urban area, household size, whether female heads the household, mother's and fathers' number of years of education, and 1992 region-level variables, which are the poverty rate and the rate of landlessness among rural households in the region. The instruments are migration rates from Vietnam's 7 regions coming from the 1992 survey and the interaction of this variable and the fraction of female in the households. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.6. OLS and IV Estimates of the Impact of Remittances on Household Entrepreneurial Activities and Household Expenditures

	Business Profit			Value of Business Equipment		
	OLS		IV	OLS		IV
	(1)	(2)	(3)	(4)	(5)	(6)
Total of Remittances	0.0814 (0.0507)	0.0924* (0.0519)	2.3921*** (0.8227)	0.3772** (0.1738)	0.4122** (0.1806)	1.3991 (1.1779)
Fraction received by Female Members		-8.6563 (6.0839)	-61.2129 (68.7697)		-27.0782* (16.3704)	46.8146 (127.2398)
F-Stats from Hausman Test			8.1700			0.7000
p-value from Hausman Test			0.0003			0.4986
N	5,998	5,998	5,998	5,851	5,851	5,851
	Education Expenditures as a Share of Total Expenditure			Health Expenditure as a Share of Total Expenditure		
Total of Remittances	0.0000 (0.0000)	0.0000 (0.0000)	0.0019*** (0.0006)	0.0001** (0.0000)	0.0000 (0.0000)	0.0000 (0.0004)
Fraction received by Female Members		-0.0003 (0.0025)	-0.0570 (0.0425)		0.0276*** (0.0038)	0.1098*** (0.0270)
F-Stats from Hausman Test			18.6400			7.2600
p-value from Hausman Test			0.0000			0.0007
N	5,998	5,998	5,998	5,998	5,998	5,998
	Food Expenditures as a Share of Total Expenditure					
Total of Remittances	-0.0004*** (0.0000)	-0.0004*** (0.0000)	-0.0016* (0.0009)			
Fraction received by Female Members		-0.0201*** (0.0054)	-0.0437 (0.0506)			
F-Stats from Hausman Test			2.0200			
p-value from Hausman Test			0.1324			
N	5,998	5,998	5,998			

Notes: Robust standard errors are reported in parentheses. All specifications contain whether household is in an urban area, household size, whether female heads the household, mother's and fathers' number of years of education, and 1992 region-level variables, which are the poverty rate and the rate of landlessness among rural households in the region. The instruments are migration rates from Vietnam's 7 regions coming from the 1992 survey and the interaction of this variable and the fraction of female in the households. *** p<0.01, ** p<0.05, * p<0.1.

Bibliography

- Abdulkadiroglu, Atila, Joshua Angrist, and Parag Pathak. 2011. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools". NBER Working Paper No. 17264.
- Abedi, Jamal, and Carol Lord. 2001. "The Language Factor in Mathematics Tests". *Applied Measurement in Education*, 14 (3): 219 – 34.
- Acosta, Pablo. 2006. "Labor Supply, School Attendance, and Remittances from International Migration: the Case of El Salvador". World Bank Policy Research Working Paper No. 3903.
- Acosta, Pablo, Cesar Calderon, Pablo Fajnzylber and Humberto Lopez. 2008. "What is the Impact of International Remittances on Poverty and Inequality in Latin America". *World Development*, 36(1): 89 – 114.
- Adams and Page. 2005. "Do International Migration and Remittances Reduce Poverty in Developing Countries?". *World Development*, 33(10): 1645 – 1669.
- Aizer, Anna. 2008. "Peer Effects and Human Capital Accumulation: The Externalities of ADD". National Bureau of Economic Research Working Paper No. 14354.
- Akresh, Richard and Ilana Redstone Akresh. 2011. "Using Achievement Tests to Measure Language Assimilation and Language Bias among the Children of Immigrants". *Journal of Human Resources*, 46(3): 647 – 67.
- Albus, Debra, Martha Thurlow, Kristin Liu, and John Bielinski. 2005. "Reading Test Performance of English-Language Learners Using an English Dictionary". *The Journal of Educational Research*, 98 (4): 245 – 54.
- Amuedo-Dorantes, Catalina and Susan Pozo. 2009. "New Evidence on the Role of Remittances on Health Care Expenditures by Mexican households". IZA Working Paper No. 4617.

- Angrist, Joshua, Aimee Chin, and Ricardo Godoy. 2008. "Is Spanish-only Schooling Responsible for the Puerto Rican Language Gap?". *Journal of Development Economics*, 85 (1-2): 105 – 28.
- Angrist, Joshua, and Kevin Lang. 2004. "Does Schooling Integration Generate Peer Effects? Evidence from Boston's Metco Program". *American Economic Review*, 94 (5): 1613 – 34.
- Argys, Laura, Daniel Rees, and Dominic Brewer. 1996. "Detracking America's Schools: Equity at Zero Cost?". *Journal of Policy Analysis and Management*, 15(4): 623 – 45.
- August, Diane, and Timothy Shanahan. 2006. *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum.
- Baker, Keith, and Adriana de Kanter. 1981. *Effectiveness of Bilingual Education: A Review of the Literature*. Washington, DC: U.S. Department of Education.
- Barecca, Alan I., Melanie Guldi, Jason M. Lindo, and Greg R. Waddel. 2010. "Running and Jumping Variables in Regression Discontinuity Designs". University of Oregon, mimeo.
- Betts, Julian, and Jaimie Shkolnik. 2000. "The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools". *Economics of Education Review*, 19(1): 1 – 15.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?". *Quarterly Journal of Economics*, 119(1): 249 – 275.
- Bhatt, Rachana. Forthcoming. "A Review of Gifted and Talented Education in the U.S." *Education Finance and Policy*, 6(4), 557 – 82.
- . 2010. "The Impacts of Gifted and Talented Education". mimeo.
- Bifulco, Robert, Jason Fletcher, and Stephen Ross. Forthcoming. "The Effect of Classmate Characteristics on Individual Outcomes: Evidence from the ADD Health". *American Economic Journal: Economic Policy*.

- Black, Sandra, Paul Devereux, and Kjell Salvanes. 2010. "Under Pressure? The Effect of Peers on Outcomes of Young Adults". National Bureau of Economic Research Working Paper No. 16004.
- Borraz, Fernando. 2005. "Assessing the Impact of Remittances on Schooling: The Mexican Experience". *Berkeley Electronic Press*.
- Card, David, and Alan Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States". *Journal of Political Economy*, 100(1): 1 – 40.
- Carhill, Avary, Carola Suarez-Orozco, and Mariela Paez. 2008. "Explaining English Language Proficiency Among Adolescent Immigrant Students". *American Educational Research Journal*, 45 (40): 1155 – 79.
- Carrell, Scott, and Mark Hoekstra. 2010. "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids". *American Economic Journal: Applied Economics*, 2 (1): 211 – 28.
- Cicala, Steve, and Roland Fryer. 2011. "A Roy Model of Social Interactions". NBER Working Paper No. 16880.
- Clark, Daman. 2010. "Selective Schools and Academic Achievement". the B.E. Journal of Economic Policy Analysis and Policy, 10, (Advances), Article 9.
- Clewell, Chu Beatriz, Michael Fix, and Jorge Ruiz-de-Velasco, 2000. "Overlooked and Underserved: Immigrant Children in U.S. Secondary Schools". Washington, D.C.: The Urban Institute.
- Cox-Edwards, Alejandra and Manuelita Ureta. 2003. "International Migration, Remittances and Schooling: Evidence from El Salvador". *Journal of Development Economics*, 72(2): 429 – 461.
- Davis, Billie, John Engberg, Dennis N. Eppe, Holger Sieg, and Ron Zimmer. 2010. "Evaluating

- the Gifted Program of an Urban School District using a Modified Regression Discontinuity Design”. NBER Working Paper No. 16414.
- Davis, James A. 1966. “The Campus a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men”. *American Journal of Sociology*, 72(July): 17 – 31.
- Dang, Nguyen Anh, Tacoli, C., and Hoang Xuan Thanh. 2003 “Migration in Vietnam: A Review of Information on Current Trends and Patterns, and Their Policy Implications”. A Paper presented at the Regional Conference on Migration, Development and Pro-Poor Policy Choices in Asia, Dhaka, Bangladesh, 22-24 June 2003.
- Dobbie, Will and Roland G. Fryer. 2011. “Exam High Schools and Academic Achievement: Evidence from New York City”. NBER Working Paper No. 17286.
- Duflo, Esther. “Grandmothers and Granddaughters: Old Age Pension on Child and Intra-Household Allocation in South Africa”. *World Bank Economic Review*, 17(1) 2003: 1 – 25.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya”. *American Economic Review*, 101(5): 1739 – 74.
- Engberg, John, Dennis Epple, Jason Imbrogno, Holger Sieg, and Ron Zimmer. 2010. “Evaluation Education Programs that Have Lotteried Admission and Selective Attrition”. University of Pennsylvania, mimeo.
- Epple, Dennis, Elizabeth Newton, and Richard Romano. 2002. “Ability Tracking, School Competition, and the Distribution of Educational Benefits”. *Journal of Public Economics*, 83(1): 1 – 48.
- Figlio, David. 2005. “Boys Named Sue: Disruptive Children and Their Peers”. National Bureau of Economic Research Working Paper No. 11277.

- Figlio, David, and Marianne Page. 2002. "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Equality?". *Journal of Urban Economics*, 69(1): 201 – 209.
- Fine, Michelle. 1991. *Framing Dropouts: Notes on the Politics of an Urban Public High School*. SUNY Series, Teacher Empowerment and School Reform. Albany, NY: State University of New York Press.
- Friesen, Jane, and Brian Krauth. 2011. "Ethnic Enclaves in the Classroom". *Labour Economics*, 18 (5): 656 – 63.
- Greene, Jay. 1998. "A Meta-Analysis of the Effectiveness of Bilingual Education" (unpublished manuscript).
- Grissom, James, and Lorrie Shepard. 1989. "Repeating and Dropping Out of School". In *Flunking Grades: Research and Policies on Retention*, ed. Lorrie Shepard and Mary Smith, 34 – 64. London: The Falmer Press.
- Gould Eric, Victor Lavy, and M. Daniele Paserman. 2009. "Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence". *Economic Journal*, 119 (540): 1243 – 69.
- Guzman, Juan Carlos, Andrew Morrison and Mirja Sjoblom. 2008. "The Impact of Remittances and Gender on Household Expenditure Patterns: Evidence from Ghana" in eds., Andrew Morrison, Maurice Schiff and Mirja Sjoblom, *The International Migration of Women*, pp. 125 – 152, World Bank and Palgrave MacMillan.
- Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design". *Econometrica*, 69(1): 201 – 209.
- Hallman, Kelly. 2000. "Mother-Father Resource Control, Marriage Payments, and Girl-Boy Health in Rural Bangladesh". FCND Discussion Paper No. 93.
- Hanson, Gordon and Christopher Woodruff. 2003. "Emigration and Educational Attainment in

- Mexico”. University of San Diego, mimeo.
- Hanushek, Eric, John Kain, Jacob Markman, and Steven Rivkin. 2003. “Does Peer Ability Affect Student Achievement?”. *Journal of Applied Econometrics*, 18 (5): 527 – 44.
- Hanushek, Eric, John Kain, and Steven Rivkin. 2009. “New Evidence about *Brown v. Board of Education*: The Complex Effects of School Racial Composition on Achievement”. *Journal of Labor Economics*, 27 (3): 349 – 83.
- Hildebrandt, Nicole, and David McKenzie. 2005. “The Effects of Migration on Child Health in Mexico”. World Bank Policy Research Working Paper 3573.
- Koekstra, Mark. 2009. “The Effect of Attending the Flagship University on Earnings: A Discontinuity Based Approach”. *Review of Economics and Statistics*, 91(4): 717 – 24.
- Hoxby, Caroline. 2000. “Peer Effects in the Classroom: Learning from Gender and Race Variation”. National Bureau of Economic Research Working Paper No. 7867.
- Hoxby, Caroline, and Gretchen Weingarth. 2006. “Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects”. Paper presented at the 2006 American Economics Association Annual Meeting.
http://www.aeaweb.org/annual_mtg_papers/2006/0108_1300_0803.pdf.
- Imberman, Scott, Adriana Kugler, and Bruce Sacerdote. Forthcoming. “Katrina’s Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees”. *American Economic Review*.
- Jackson, Kirabo, “Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignment in Trinidad and Tobago”. *The Economic Journal*, 120(549): 1399 – 1429.
- Jacob, Brian, and Lars Lefgren. 2009. “The Effect of Grade Retention on High School Completion”. *American Economic Journal: Applied Economics*, 1 (3): 33 – 58.

- Kane, Thomas J. and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation". NBER Working Paper No. 14607.
- Kulik, Chen-lin, and James Kulik. 1997. "Ability Grouping" in Handbook of Gifted Education, edited by Nicholas Colangelo and Gary Davis, 230 – 242. Boston: Allyn and Bacon.
- Lavy, Victor, Daniele Paserman, and Analia Schlosser. 2008. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in High and Low Achievers in the Classroom". NBER Working Paper No. 14415.
- Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School". *American Economic Journal: Applied Economics*, 3 (2): 1 – 33.
- Lazear, Edward. 1999. "Culture and Language". *Journal of Political Economy*, 107 (6): S95 – S126.
- Lee, David, and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics". *Journal of Economic Literature*, 48(2): 281 – 355.
- Loveless, Tom, Steve Farkas, and Ann Duffett. 2008. "High Achieving Students in the Era of NCLB". Thomas B. Fordham Institute Report.
- Lopez-Cordova. 2005. "Globalization, Migration and Development: The Role of Mexican Migrant Remittances". *Economia*, 6(1): 217 – 256.
- Lundberg, Shelly, Robert Pollak, and Terence Wales. "Do Husbands and Wives Pool Their Resources? Evidence from the United Kingdom Child Benefit". *Journal of Human Resource*, 32(3): 463 – 480.
- Master, Benjamin, Susanna Loeb, Camille Whitney, and James Wyckoff. 2012. "Different Skills? Identifying Differentially Effective Teachers of English Language Learners". AEFPP 2012 Conference paper.
- Matsudaira, Jordan. 2005. "Sinking or Swimming? Evaluating the Impact of English Immersion versus Bilingual Education". Working paper.

- McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test". *Journal of Econometrics*, 142(2): 698 – 714.
- McKenzie, David and John Gibson. 2010. "The Development Impact of a Best Practice Seasonal Worker Policy". World Bank Policy Research Working Paper No. 5488.
- McKenzie, David, John Gibson and Steven Stillman. Forthcoming. "The Impacts of International Migration on Remaining Household members: Omnibus Results from a Migration Lottery Program". *Review of Economics and Statistics*.
- McKenzie, David, John Gibson and Steven Stillman. 2010. "How Important is Selection? Experimental Versus Non-experimental Measures of the Income gains from Migration". *Journal of the European Economic Association*, 8(4): 913 – 945.
- McKenzie, David and Hillel Rapoport. 2010. "Can Migration Reduce Educational Attainment? Evidence from Mexico". *Journal of Population Economics*, 1 – 28.
- Murphy, Patrick Ryan. 2009. "Essays on Gifted Education's Impact on Student Achievement". Doctoral dissertation, Florida State University.
- Neal, Derek, and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability". *Review of Economics and Statistics*, 92(2): 263 – 283.
- Niimi, Yoko, Thai Hung Pham, and Barry Reilly. 2009. "Determinants of Remittances: Recent Evidence Using Data on Internal Migrants in Vietnam". *Asian Economic Journal*, 23(1): 19 – 39.
- Ohinata, Asako, and Jan C. van Ours. 2011. "How Immigrant Children Affect the Academic Achievement of Native Dutch Children". IZA Discussion Paper No. 6212.
- Pop-Eleches, Cristian, and Miguel Urquiola. 2011. "Going to a Better School: Effects and Behavior Responses". NBER Working Paper No. 16886.
- Quisumbing, Agnes and John Maluccio. 2003. "Resources at Marriage and Intrahousehold

- Allocation: Evidence from Bangladesh, Ethiopia, Indonesia and South Africa”. *Oxford Bulletin of Economics and Statistics*, 65(3): 283 – 327.
- Reback, Randall. 2008. “Teaching to the Rating School Accountability and the Distribution of Student Achievement”. *Journal of Public Economics*, 92(5-6): 1394 – 1415.
- Roderick, Melissa. 1994. “Grade Retention and School Dropout: Investigating the Association”. *American Educational Research Journal*, 31 (4): 729 – 59.
- Rossell, Christine, and Keith Baker. 1996. “The Educational Effectiveness of Bilingual Education”. *Research in the Teaching of English*, 30 (1): 7 – 74.
- Rothstein, Jesse. 2010. “Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement”. *Quarterly Journal of Economics*, 125(1): 175 – 214.
- Rumberger, Russell. 1987. “High School Dropouts: A Review of Issues and Evidence”. *Review of Educational Research*, 57 (2): 101 – 21.
- Small, Ivan, Truong thi Kim Chuyen, and Diep Vuong, “Diaspora Philanthropy in Vietnam”, in *Diaspora Giving: An Agent of Change in Asia Pacific Communities?*, edited by Paula Johnson. Manila: Asia Pacific Philanthropy Consortium, 2008. 251-283.
- Suarez-Orozco, Carola, and Marcelo Suarez-Orozco. 2001. *Children of Immigration*. Cambridge, MA: Harvard University Press.
- Thomas, Duncan. 1990. “Intra-household Resource Allocation: an Inferential Approach”. *Journal of Human Resources*, 25(4): 635 – 664.
- Thomas, Duncan. 1994. “Like Father, Like son; Like Mother Like Daughter: Parental Resources and Child Height”. *Journal of Human Resources*, 29(4): 950 – 988.
- Wang, Jia, and Pete Goldschmidt. 1999. “Opportunity to Learn, Language Proficiency, and Immigrant Status Effects on Mathematics Achievement”. *The Journal of Educational Research*, 93 (2): 101 – 11.

- Willoughby, Louisa. 2009. "Language Choice in Multilingual Peer Groups: Insights from an Australian High School". *Journal of Multilingual and Multicultural Development*, 30 (5): 421 – 35.
- World Bank. 1999. "Vietnam Attacking Poverty". Vietnam Development Report 2000, Poverty Reduction and Economics Management Unit, East Asia and Pacific Region. World Bank Vietnam Office, Hanoi
- Yang, Dean. 2008. "International Migration, Remittances, and Household Investment: Evidence from Philippine Migrants' Exchange Rate Shocks". *Economic Journal*, 118(528): 591 – 630.
- Yang, Dean, 2011. "Migrant Remittances". *Journal of Economic Perspectives*, 25(3): 129 – 152.

